PRINCETON UNIVERSITY LIBRARY
APR 9 1954
RECEIVED

OXFORD ECONOMIC PAPERS

(NEW SERIES)

VOLUME 6 FEBRUARY 1954 NUMBER 1

Editorial Board: T. WILSON (*General Editor*), J. R. HICKS, A. BRIGGS, C. M. KENNEDY, G. B. RICHARDSON (*Secretary*)

CONTENTS

The Trade Cycle in Britain, 1790–1850. By R. C. O. MATTHEWS . 1

Notes on Excess Capacity. By M. E. PAUL 33

The Process of Imperfect Competition. By J. R. HICKS . . . 41

Convertibility. By J. R. SARGENT 55

Some Economic Aspects of Compulsory Trade Unionism. By V. L. ALLEN 69

An Experiment in Commodity Control—The International Wheat Agreement, 1949–53. By C. D. HARBURY 82

An Alternative Proof of a Theorem in Welfare Economics. By CHARLES KENNEDY 98

OXFORD
AT THE CLARENDON PRESS
1954

OXFORD ECONOMIC PAPERS

(NEW SERIES)

OXFORD ECONOMIC PAPERS, of which eight numbers appeared at irregular intervals between 1938 and 1947, became a periodical in 1949 and is now published three times a year. The pre-paid subscription price is 25*s*., post free, and the price of a single copy 12*s*. net. Orders may be placed with any bookseller or sent direct to the Publisher.

THE JOURNAL consists, in the main, of articles by Oxford authors, but the editors will also be glad to consider contributions from elsewhere. As in the past, articles of considerable length, which embody the results of substantial research, are included from time to time. The editors also publish occasional articles on Economic History, Public Administration, and Scientific Method, provided these articles are likely to be of general interest to economists. Contributors receive 50 free offprints.

PAPERS should be sent to the General Editor, c/o The Clarendon Press, Oxford. All correspondence other than that dealing with contributions should be addressed to the Publisher:

GEOFFREY CUMBERLEGE, OXFORD UNIVERSITY PRESS
AMEN HOUSE, LONDON, E.C. 4

OXFORD ECONOMIC PAPERS

NEW SERIES
VOLUME VI
1954

OXFORD
AT THE CLARENDON PRESS
1954

Oxford University Press, Amen House, London E.C.4

GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS KARACHI CAPE TOWN IBADAN

Geoffrey Cumberlege, Publisher to the University

PRINTED IN GREAT BRITAIN
AT THE UNIVERSITY PRESS, OXFORD
BY CHARLES BATEY, PRINTER TO THE UNIVERSITY

CONTENTS

VOLUME 6

Number 1, February 1954

THE TRADE CYCLE IN BRITAIN, 1790–1850. *By* R. C. O. MATTHEWS . 1

NOTES ON EXCESS CAPACITY. *By* M. E. PAUL 33

THE PROCESS OF IMPERFECT COMPETITION. *By* J. R. HICKS . 41

CONVERTIBILITY. *By* J. R. SARGENT 55

SOME ECONOMIC ASPECTS OF COMPULSORY TRADE UNIONISM. *By* V. L. ALLEN 69

AN EXPERIMENT IN COMMODITY CONTROL—THE INTERNATIONAL WHEAT AGREEMENT, 1949–53. *By* C. D. HARBURY . . 82

AN ALTERNATIVE PROOF OF A THEOREM IN WELFARE ECONOMICS. *By* CHARLES KENNEDY 98

Number 2, June 1954

VARIATIONS IN THE RATE OF ECONOMIC GROWTH IN THE U.S.A., 1869–1939. *By* B. WEBBER and S. J. HANDFIELD-JONES . . 101

AN INTERPRETATION OF ECONOMIC BACKWARDNESS. *By* H. MYINT 132

RECONSIDERATION OF A GROUP WELFARE INDEX: A REJOINDER ON MARGINAL PREFERENCE. *By* J. ROTHENBERG . 164

THE MACMILLAN GAP 1931–53. *By* R. FROST 181

Number 3, September 1954

PERIOD ANALYSIS AND INVENTORY CYCLES. *By* R. NURKSE . 203

THE PRICE OF FUEL. *By* D. L. MUNBY 226

THE DOLLAR CRISIS REVISITED. *By* T. BALOGH 243

INVESTMENT DECISIONS, TRADE CYCLE, AND TREND. *By* A. J. YOUNGSON 285

SOME ELEMENTARY THEORY ABOUT ACCUMULATION. *By* D. B. BUTT 306

Editorial Board

T. WILSON (*General Editor*) J. R. HICKS, A. BRIGGS,
C. M. KENNEDY, G. B. RICHARDSON (*Secretary*)

THIS SERIAL PUBLICATION is intended primarily as a channel for the publication of articles by Oxford authors. Contributions from elsewhere will, however, be considered, and should be sent to the General Editor. While most of the articles will be concerned with Economics, the editors also hope to publish occasional articles on Economic History and Public Administration provided these are likely to be of general interest to economists.

THE TRADE CYCLE IN BRITAIN, 1790–1850

By R. C. O. MATTHEWS

I

WITH the long delayed publication of the results of the work on British business cycles from 1790 to 1850 carried out under the auspices of the Columbia University Council for Research in the Social Sciences,[1] students are provided with what is perhaps the fullest picture that has yet been made available of the evidence relating to fluctuations in the British economy in any period of its history. The work was carried out under the direction of the late A. D. Gayer between 1936 and 1941. Various obstacles have caused its publication to be delayed till now, and in an addendum to the preface Professor Rostow and Mrs. Schwartz, who were Professor Gayer's chief collaborators, briefly review contributions to the field that have appeared since the work was completed.

The book falls into four clearly defined divisions. The first (Vol. I, Part I) is a chronological survey of the period. The second (Vol. I, Parts II and III) presents and discusses two new groups of index numbers. The first of these groups relates to share prices. This section of the work has already been published in part.[2] The indices of share prices are accompanied by a chapter on their institutional background, which contains material on the history of the various types of joint-stock enterprise launched during the period. This chapter has been condensed from a much longer original text which has been microfilmed but is not to be published in full. The second group of index numbers relates to commodity prices. These indices are likely to become the standard ones for use by students in the future, superseding Silberling's index, of which the authors have some pertinent criticisms to make. The third division of the work (Vol. II, Part I) is the one which is likely to have the greatest number of readers. It contains a general theoretical analysis of the nature and causes of the fluctuations chronicled in Vol. I, Part I, followed by a briefer discussion of trend movements over the period. Finally, in Part II of Vol. II, the chief statistical series available are charted and subjected to measures of average cyclical behaviour according to the technique of the National Bureau of Economic Research. The results indicated by these measures are discussed in a running commentary accompanying the charts. The statistical tables

[1] *The Growth and Fluctuation of the British Economy, 1790–1850*, by A. D. Gayer, W. W. Rostow, and A. J. Schwartz, assisted by I. Frank. Oxford University Press, 1953. Two volumes, pp. xxxvi+xii+1108, 84*s*.

[2] A. D. Gayer, A. Jacobson, and I. Frank, 'British Share Prices, 1811–1850', *Review of Economic Statistics*, 1940, pp. 78–93.

HB 836 Ser. 2 v. 6 no. 1 (1954/55)

on which this part of the work is based are not included, but form part of the ancillary microfilmed material.

The principal purpose of the present article is to describe and criticize in as constructive a manner as possible the leading theses advanced in Part I of Vol. II. These theses are reflected at many points in the discussion of the average measures of cyclical behaviour, and they also influence both the arrangement and the substance of the narrative section of the work. Despite its title, the book is much more concerned with fluctuations than with growth, and we shall pass over without discussion the brief but interesting and, to the present writer, largely acceptable suggestions about trend movements over the period that are put forward in Chapters IV and V of Vol. II, Part I.

Both the narrative section and the theoretical analysis are based on the distinction drawn by the authors between major and minor cycles. The application of this distinction to the period 1790–1850 has become familiar from Professor Rostow's *British Economy of the Nineteenth Century*, which was, of course, written after the present work had been completed. The major cycles are (trough-peak-trough) 1788–92–3, 1797–1800–1, 1808–10–11, 1816–18–19, 1819–25–6, 1832–36–7, and 1842–5–8; the minor cycles are 1793–6–7, 1801–2–3, 1803–6–8, 1811–15–16, 1826–8–9, 1829–31–2, and 1837–9–42.[1] Major and minor cycles thus usually occur alternately, and the major cycles are on the average of longer duration. The contraction phases of both major and minor cycles in most cases last for only one year.

In accordance with the customary National Bureau procedure, the dates given for turning-points are intended to show turning-points in 'general business activity', as emerging from the consensus of statistical data, rather than turning-points in any particular magnitude such as national income. General business activity is not defined. The authors suggest, however, that the ambiguity in this respect is more apparent than real, on the grounds that the consensus of the chief statistical series shows that fluctuations were both pervasive and persistent. This view is, of course, not without its difficulties, and some of these will be discussed presently.

The distinction between major and minor cycles depends partly on the amplitude of the fluctuations involved; but the main basis of the distinction is that major cycles are those which included a period of substantial domestic investment. The authors' thesis is that the minor cycles arose chiefly from fluctuations in exports, whereas the expansion phases of major cycles, although initiated by increases in exports, depended at their later stages on domestic investment. This pattern they consider to have

[1] For certain purposes the years 1797–1803 are treated as a single major cycle. A further minor cycle 1846–7–8 is recognized when dealing with monthly as distinct from annual data. The years 1819–21 also partake of some of the characteristics of a minor cycle.

persisted in essentials even during the war period, although at that time capital export to some extent took the place of domestic investment.

Space does not permit discussion of the many interesting points that arise out of the narrative section of the book and out of the average measures of cyclical behaviour. The statistical and other material which the authors have passed in review is very extensive, and some of it is little known. This does not mean that there are not some sources which might have been more fully used,[1] and specialists in particular fields will find a fair number of minor misstatements. The presentation of some of the statistical material is also open to criticism.[2] The narrative is arranged in standardized form, the treatment of each cyclical period consisting of six sections, dealing respectively with prices, foreign trade, investment, industry and agriculture, finance, and labour. This arrangement greatly facilitates reference, but it tends to hamper the development of any sustained argument, and on occasions threatens to degenerate into a scrapbook. Criticism is partially disarmed by the authors' admission (p. 5) that

[1] In particular, many of the relevant Parliamentary Papers, especially those dealing with money and banking, appear not to have been very carefully worked through. No use is made of the weekly *Circular to Bankers*, which is a mine of information for the second quarter of the of the nineteenth century.

[2] Certain particular faults may be enumerated for the benefit of those who wish to use the volumes for purposes of reference. (*a*) The figures given in the narrative section as those of the Official values of imports, exports, and re-exports are, in fact, not the Official values but Schlote's calculations of volume, which are computed on a different basis. (W. Schlote, *Entwicklung und Strukturwandlungen des englischen Aussenhandels vor 1700 bis zur Gegenwart*, 1938; English edition, *British Overseas Trade from 1700 to the 1930's*, translated by W. O. Henderson and W. H. Chaloner, 1952.) These are also the figures that are charted and analysed in Part II of Vol. II. Schlote's estimates and the Official values behave broadly similarly, but their movements are not identical. In the case of imports there are 4 years in which the direction of change of the two are different, in the case of exports 2 years; and there are also a good many differences in the relative magnitudes of movements as between different years. In the case of exports, moreover, the difference of base causes the Official values to exceed Schlote's index throughout by a factor of about two. (See also below, p. 20, n. 1.) (*b*) The figures given for consumption of raw cotton—regarded as an index of production of cotton textiles—are derived from a variety of sources and are, as the authors remark, unsatisfactory. 'It is claimed that the series was derived by correcting net cotton imports for changes in stocks; but the series seems to partake of the nature of inventory rather than production variations. . . . It is virtually identical, in its short-period movements, with total cotton imports. It is clear from the history of the period that, for example, cotton production did not fluctuate from 1819 to 1826 in the highly irregular manner that the cotton consumption series would indicate' (pp. 700–1). The figure in T. Ellison's *Cotton Trade of Great Britain* (1886), which cover the years after 1811, are much more plausible than those which the authors have used, and might with advantage have been given instead. (*c*) The so-called 'Kondratieff index of money wages in the textile industry' is merely G. H. Wood's index of money wages in the cotton industry, adjusted during the Bank Restriction period by a gold conversion factor that is likely to mislead the unwary. Wood's figures (*History of Wages in the Cotton Trade*, 1910) are themselves a weighted average of two series relating respectively to the wages of factory operatives and of hand-loom weavers, and the diminishing weight accorded to hand-loom weavers' wages as the period advances contributes substantially to the upward trend of the composite series. It would surely have been better to have given separately Wood's two series, the movements of which are largely dissimilar.

the narrative has deliberately been kept on a more or less 'journalistic' plane so as to make it independent of the more speculative hypotheses contained in Part I of Vol. II. The result is, however, that many of the most interesting and difficult questions about particular periods go unanswered or even unasked.

The measures of cyclical behaviour in Vol. II, Part II, follow the standard National Bureau technique, as described at length in A. F. Burns and W. C. Mitchell's *Measuring Business Cycles* and conveniently summarized in Chapter 3 of Mitchell's posthumous *What Happens during Business Cycles*.[1] This technique has been subjected to various criticisms in the past, and the results it yields certainly require to be handled with great caution; but much that is interesting emerges. For certain selected series separate measures are given for major and minor cycles, and it would have been interesting to see this done on a larger scale. Since the minor cycles are to be regarded from certain points of view as part of the contraction phase of the preceding major cycles, it might also have been valuable for some measures to have been given on the basis of one major plus one minor cycle as the cyclical unit—as it would have been to have had separate measures of average cyclical behaviour during the war and during the rest of the period.

The arrangement of the remainder of this article is as follows. Section II contains a general discussion of the behaviour of investment, which, as has been said, provides the criterion of major cycles as defined by the authors. Section III concerns certain offsets to investment which were of importance especially, though not exclusively, during the war period. Section IV is devoted to foreign trade. Foremost among the issues which we are passing over for reasons of space are those that relate to monetary aspects of fluctuations and their connexion with the balance of payments.

II

It may be well to begin by describing briefly the evidence available on investment. Much the most important statistical series is that recording the number of bricks produced annually, a series which starts in 1785 and does not cease until the repeal of the duty on bricks in 1850. From 1829 to 1849 (and to a limited degree for part of the earlier period) figures of brick production are available on a regional basis, but these regional series are not used in the present work. Further evidence on the level of building activity comes from the figures on timber imports, which move in a pattern broadly similar to that of the brick index. Almost the only other continuous and unambiguous series relating to investment is that recording

[1] A brief description of part of the National Bureau's technique of measurement is given below, p. 18, n. 2.

the tonnage of new ships constructed annually. The remaining evidence, though important, is heterogeneous. We have particulars of the annual number of private Acts of Parliament sanctioning the construction of railways, canals, and various other public utilities, and the amount of capital authorized in each case; but the proportion of the sums authorized that was actually spent and the timing of this expenditure are not always easy to ascertain. Much the same applies to the inferences that may be drawn about investment in agriculture from the number of bills of enclosure. Evidence on the number of joint-stock company prospectuses issued exists for certain years and may be compared with the figures of private Acts of Parliament. Finally we have a good deal of evidence which is non-continuous or non-quantitative or both on investment in manufacturing industry and other branches of the economy. It will be seen that the evidence is thus very heterogeneous and, such as it is, is biased in favour of that part of investment consisting of construction.

The account given by the authors of the time pattern of investment and the designation of particular cycles as major and minor are based chiefly on the brick index, but the other measures of investment quoted give broadly consistent results. Brick production was low or declining during each of the minor cycles, with the exception of 1837–42, and, less clearly, 1801–8 (years during which the cyclical pattern as a whole is rather confused). It rose during every major cycle, with a marked tendency to lag at both upper and lower turning-points behind other indicators of the level of activity. In the case of contraction phases limited to 1 year in duration, this lag sometimes appears as a net rise during the contraction phase. The lag at the lower turning-point is one of the considerations that gives rise to the thesis that major expansions were *initiated* by causes outside the sphere of investment, such as increases in exports.

Most of the quantitative evidence on investment relates, as has been seen, to building and public utilities. The theoretical treatment of fluctuations in investment in Vol. II, Part I, of the work is, however, cast mainly in terms of manufacturing industry. This theory will now be described.

It is based in essence on two concepts: the gestation period of investment and the secular growth of demand. Let us start at the beginning of the expansion phase of a major cycle. There is excess capacity, and the time is not yet ripe for a renewal of domestic investment. A lapse of time is necessary before investment will rise, partly in order for demand to catch up with the existing level of capacity, partly because a decent level of profits will have to be maintained for some while before entrepreneurs can be persuaded to take a sufficiently optimistic view of the future to engage in long-term commitments, and partly because investment cannot be done unless entrepreneurs and investors have accumulated sufficient capital to

finance it.[1] The catching-up process is going on the whole time on account of the secular growth of demand; and the rise in profits, which is the second necessary condition, is brought about by the increase in exports which is the characteristic of the early stages of major expansions. (Whether the authors regard this increase in exports as something distinct from the secular growth in demand or merely as part of it is not entirely clear.[2]) Other favourable factors may also make a contribution, such as good harvests, which were in fact present at the beginning of almost all major expansions. At length the various conditions are satisfied which are necessary for a large amount of investment to be set on foot, and the expansion moves into its second and culminating stage. Now additions to capacity must, for technical reasons, be made in fairly large indivisible units, and the investment involved has an appreciable gestation period. All firms are operating under the influence of broadly similar stimuli, and the working of competition thus causes capacity to be expanded beyond that justified by the current level of demand. It is this which keeps investment low in the years after the major peak. However, this is not what brings about the recession in the first place. The authors say:

> The upper turning-point came, in fact, before all the acts of new investment were complete; and the level of employment in several minor cycles was supported by the completion of projects inaugurated in the preceding boom. Our theory of the upper turning-point does not depend, then, on a fall of profits brought about by the actual introduction of new equipment; although knowledge of the scale of investment in certain lines could, and did, affect expectations of profit from further investment, even though the investments already undertaken were not yet in full operation. *Our view is dependent on this latter factor combined with the consequences for costs of relatively full employment, partially created by the domestic industrial investment. . . .*[3]

The minor cycle that follows on the major cycle is essentially part of the contraction phase of the latter, and its upswing, such as it is, is based on factors other than domestic investment. When the minor contraction phase is completed, a new major expansion begins, and we are back to the beginning again.

This outline statement of the theory is capable of modification in various respects to suit individual cycles. It will have been observed, for example, that fluctuations in the level of final demand are not essential to the explanation of fluctuations in investment just described; but the authors are perfectly willing to concede that some sort of multiplier-accelerator interaction was at work as well and that the high level of consumers' demand

[1] The most obvious application of this is to investment financed out of ploughed back profits; but it is relevant also to projects like railways, financed by public subscription, since investors commonly had to go on paying calls on shares for quite a long time after buying them, and so would not be ready for some while after one boom to assume fresh obligations and so help start another.

[2] See below, p. 28.

[3] p. 557. Italics added.

in the boom helped to account for the high level of investment then, and conversely in the slump. Likewise although monetary stringency is not regarded as the basic cause of the upper turning-point, it is freely admitted that rises in interest rates were a regular feature of these periods and that they may have contributed to a greater or less degree to bringing the boom to an end on various occasions.

Apart from possible disagreements about matters of emphasis, on which the authors are quite undogmatic, certain more fundamental difficulties about this theory suggest themselves.

1. Reference to the gestation period makes it easy to understand why investment should have lagged behind the decisions giving rise to it. But the fact that investment (as measured chiefly but not exclusively by the brick index) should have continued to rise in the early stages of the recession and should even sometimes have risen during the whole of the contraction phase of the major cycle raises a difficulty which the authors appear not to have sufficiently considered. If investment expenditure continues to rise after the peak, what is it that does decline at the beginning of the recession? What is the recession a recession of? A decline in investment decisions does not lead to a recession in the normal sense of the term until the decisions come to be translated into action. And if the recession originates outside the sphere of investment, some revision in the theory would appear to be required.

2. The thesis that investment decisions decline partly because entrepreneurs foresee that the current rate of construction will presently bring into existence an excessive amount of capacity in the industry is obviously a plausible one, even though the authors do not cite much evidence that can be considered to support it directly. That expectations should to some degree operate in this way may perhaps be taken as sufficiently obvious to make specific evidence to that effect unnecessary. (That this factor was the *decisive* one at the upper turning-point is, of course, not obvious in this sense.) But the same cannot be said of the suggestion that investment decisions were checked by rising costs attributable to the approach to full employment. As far as I have been able to discover, this suggestion as applied to 1790–1850 is a piece of pure theorizing, unsupported by any evidence whatever.[1]

The authors' treatment of this subject is also unsatisfactory on other grounds. In the first place, it may be questioned whether the concept of full employment is really a very helpful one as applied to an economy such

[1] Unless one considers as such the rise usually shown in the later stages of major expansions by the available price indices, which in any case relate mainly to the prices of food and raw materials and are not very relevant to the cost of investment.

as Britain's in the early nineteenth century, when the chief occupations were agriculture and domestic service, and the labour force of the leading manufacturing industry consisted largely of juveniles.[1] In the second place, although the term 'full employment' is often on the authors' lips, especially in the narrative section of the book, there seems to be some lack of clarity in their minds as to what is the factor of production involved. At some points they appear to have in mind the full employment of capacity—yet on their own account it is the approach to this situation which *starts* investment booms. At other times they talk as if all factors of production—labour, capacity, and even monetary facilities—were likely to approach full employment at the same time—a position not easy to defend theoretically. The best case in *a priori* terms could be made for the supposition that full employment either of capacity or of labour in the investment goods industries was the bottleneck—for example, the capital equipment of textile machinery manufacturers, or the labour force attached to the building industry. This, however, is not made clear, and in any case, as has been said, it remains at best no more than an interesting conjecture unsupported by evidence.

3. The theory is stated chiefly with reference to investment in manufacturing industry. It can, however, evidently be applied in principle to investment in such lines as housebuilding and shipbuilding. The authors do not attempt to apply it to investment in agriculture, which was an important item during the war period. The principal factor here was clearly the price of wheat, which in turn depended on harvests and the availability of foreign supplies.[2] More important, however, is the question to what extent the theory is applicable to investment in the transport

[1] An interesting example of the difficulties involved in the concept of full employment arises from some little-known statistics which the authors have unearthed from a Parliamentary Paper of 1890–1 (1890–1, xcii, pp. 522–3). These figures show the proportion of unemployed to total members of the Friendly Society of Ironfounders of England, Ireland, and Wales, beginning with the year 1831. The movements in this proportion show a very close negative conformity to the normally accepted dates of business cycles, as also do those in payments by the Society for unemployment benefit (also, surprisingly, for sick benefit). On looking up the reference one finds that it also contains figures for the total membership of the Society, which the authors do not quote. There are no doubt certain objections to taking the Society's membership as indicative of the total number of workers attached to the industry; but if they are taken as an approximate indication of this, the unemployment proportion appears on several occasions in rather a different light. A notable instance is in 1838–40. In these years the number and proportion of unemployed men were much higher than in the general peak year 1836; but the rise in membership in the meanwhile had been so great that the number of employed members of the Society was much higher also, not declining substantially till 1841. It is interesting to compare this with the rather similar picture shown by the statistics relating to active and idle blast-furnaces in the industry at the same period; see R. C. O. Matthews, *A Study in Trade-Cycle History: Economic Fluctuations in Great Britain, 1833–42*, pp. 158–9.

[2] It is possibly true that the extension of agricultural capacity undertaken in years of high prices during the war influenced prices even in the short run, and that the significance of this has often been underrated. But harvest fluctuations remained the dominant factor.

industries. Canal construction played a leading part in the boom of the 1790's; railway building was of great importance in 1836 and absolutely dominant in 1845. In the case of railways, a national network was built up from virtually nothing in the course of little more than a decade, and while the development of the canal system was a more gradual process, it was of the same innovatory nature. It follows therefore that in both cases the concept of the rate of investment being governed by the secular rate of growth of demand is inapplicable. It is most unplausible to maintain that the principal reason for the delay between the railway promotion booms of 1836 and 1845 was that time had to elapse before traffic on the older lines became sufficient to justify building new ones, and one could not even begin to apply the theory to the first railway boom. In its chief features, therefore, the theory propounded by the authors is inapplicable to these highly important branches of investment, although the subsidiary parts of the theory—those relating to finance and to confidence—remain broadly relevant.

4. There are certain special problems about the behaviour of investment during the war. These bring into some doubt the validity of the distinction between major and minor cycles as applied to that time. This question is discussed later (section III). For the present our discussion will be confined mainly to the years of peace.

In the light of these criticisms, it must be concluded that the theory in its simple form is open to serious objections. It must be said in fairness, however, that, particularly in their summing up (pp. 571–6), the authors do qualify the theory by reference to a wide range of complicating factors. Many of the points discussed in the following pages are touched on with greater or less emphasis at various stages in the book. It may be useful, however, to try to bring some of the more important factors into clearer perspective.

Perhaps the most important deficiency of the theoretical statement described is its insufficient emphasis on the speculative character of major booms in this period. This character is seen in the invariable tendency of peace-time major expansions to culminate in what contemporaries called 'manias'. These manias were closely connected with the institution of the joint-stock company, the development of which during the period left an important imprint on the character of business fluctuations. Speculation in the shares of newly formed joint-stock companies played a part in British business cycles at this period somewhat similar to that played in the United States then and later by speculation in real estate. The violent outbursts of joint-stock company promotion in 1792, 1807, 1825, 1836, and 1845 cannot be explained entirely with reference to objective calculations

of profit in the enterprises proposed. At least as important were, in the promoters' minds, the desire for commissions and profit by the sale of shares at a premium, and, in the minds of investors, the desire for quick gains by the resale of shares on a rising market. The speculative nature of the whole proceeding was enhanced by the novel and untried character of the undertakings usually canvassed. The development of a boom of this sort was essentially dependent on the creation of a general expectation that the market was rising; once under way it proceeded by its own momentum. Naturally such an expectation was in the normal way unlikely to take hold unless trade was in a generally prosperous state to start with. For this reason speculative manias usually came towards the end of a period of business expansion, and thereby contributed to the concentration of investment in the later stages of major expansions. But the basis of the mania remained largely irrational, and the nature of the schemes canvassed was governed rather by the type of enterprise in which the joint-stock principle was acceptable than by consideration of which sectors of industry had been particularly favoured by any preceding business expansion.[1] Many of the joint-stock companies so begun did not, it is true, ever lead to the creation of any physical capital; yet those that did were sufficient to account for a large part of the nation's railway and canal network, and much else besides.

The part played by pure speculation at the culmination of the upswing helps to explain many of the features of cyclical fluctuation which appear from the evidence collected by Professor Gayer and his associates, including some of the features that are not satisfactorily explained by their own simplified model. In particular it helps to explain the nature of the upper turning-point and the difficulty about the lag of the brick index and other indicators of investment at that stage. Any purely speculative movement is a highly unstable phenomenon, liable to be halted by a comparatively slight obstacle like a small rise in interest rates and incapable of avoiding collapse when once a halt has occurred. Tooke's admirable description of events in 1825 could be applied with only slight modifications to 1836 and 1845 as well:

> The recoil from speculations in loans and shares . . . was inevitable. The process by which the fall took place is simple and obvious: As regarded the schemes, a more accurate appreciation of a greater outlay, and of smaller returns, than had been before anticipated; and a limitation of the demand for investment in them, to such persons only as could afford to depend upon remote contingencies for an income, where any income was to be expected: above all, a general deficiency of means among

[1] An extreme example of this may be seen in the years 1807–8, when an outburst of a series of 'new projects of various kinds, such as canals, bridges, fire offices, breweries, distilleries and many other descriptions of joint-stock companies' (T. Tooke, *History of Prices*, i. 277–8) apparently owed its origin largely to the speculative feeling engendered by the rise in import prices brought about by the development of the war!

the subscribers to pay up the succeeding instalments, as they had relied for the most part upon a continued rise, to enable them to realise a profit before another instalment should be called for, or upon the same facility as had before existed, of raising money for the purpose at a low rate of interest. . . . A pause naturally ensued: and, under such circumstances, a pause is generally fatal to projects that do not proceed on solid grounds.[1]

Readers of Schumpeter's *Business Cycles* will be familiar with the distinction Schumpeter draws between the recession and the depression.[2] The recession is the contraction phase of the 'primary wave', and is the period during which are felt the logical consequences of the preceding boom in innovatory investment. National income declines as investment projects come to completion, and older firms are driven to bankruptcy by the lowering of prices which the expansion of capacity has brought about. The depression, which immediately follows on the recession, is the contraction phase of the 'secondary wave'. It reflects the reaction from the non-innovatory commitments which are likely to have been irrationally—or at least optimistically—undertaken during the boom and which can be supported only as long as boom conditions continue. The magnitude of the secondary contraction depends on the extent and character of these commitments and on the amount of damage done to expectations and commercial confidence by the price falls and business failures (including failures of banks) in the recession phase. Now the course of events at the upper turning-points of many of the major cycles in the period under discussion may perhaps best be summarized by saying that the speculative character of the concluding stages of the boom caused a severe secondary contraction, in Schumpeter's sense, to come before or, at least, not later than the primary contraction. The more extreme speculative manifestations of the boom—often including some elements of criminal fraud—bred their own reaction before investment had proceeded to the point where the increase in the capacity had become really damaging to profits. An important part in this secondary contraction—as in the preceding mania and subsequent minor recovery—lay in the sphere of foreign trade; this aspect of the matter will be considered later on. Falls in prices of shares and commodities led to bankruptcies and discredit, bringing with them on some occasions widespread bank failures. This was the essence of such crisis years as 1826 and 1837. The advent of this secondary contraction itself brought about a downward revision of investment decisions in sectors of the economy that were not yet feeling any great pressure on profits from previous increases in investment.

The expansion phase of the ensuing minor cycle consisted essentially of recovery from the extremes of discredit that prevailed at the trough of the

[1] T. Tooke, *History of Prices*, ii. 158–9.
[2] J. A. Schumpeter, *Business Cycles* (1939), i. 145–61.

secondary contraction. No increase in investment in fixed capital was as yet normally forthcoming, since the effects of the *primary* contraction were felt in their full force throughout the length of the minor cycle. Investment planned during the mania period continued to be carried out after the mania had come to an end, and the burden of this became plain as soon as the dust raised by the crisis had settled. The precise pattern of the minor cycle depended largely on the gestation period of the investment planned during the preceding major boom. Thus the investment preponderating in 1825 was apparently of short gestation period, and although the immediate cause of the recession was of a 'secondary' nature, it was not long before the full effects of the primary contraction were also felt. Hence a bottom was reached as early as 1826, and after some recovery in 1827 the years 1828–32 were not marked by any conspicuous further deterioration. In 1836, on the other hand, the investment projected (largely railways) was of much longer gestation period. Important categories of investment continued to rise during the minor expansion after 1837, and the full force of the primary contraction was not felt until 1841–2. The years 1820–1 were broadly comparable to 1827–32, and 1846–8 to 1837–42.

It should not, of course, be supposed that the speculative forces described above were the only or even necessarily the most important cause of fluctuations in investment. The factors affecting the level of investment were numerous and diverse, and there are many unsolved problems relating to particular industries and periods. In the case of the beginnings of the railway boom of the 1830's, I have suggested elsewhere[1] that the Schumpeterian concept of the clustering of innovations is fairly exactly applicable. Likewise the great railway boom of the 1840's can certainly not be treated exclusively in terms of speculation, although the most appropriate theoretical categories in which to describe it are not altogether obvious. In many branches of investment the interaction of an upward trend in demand and a gestation period, as described in the Columbia authors' model, was no doubt of paramount or at least substantial importance. Fluctuations in housebuilding, for example, lend themselves well to analysis in these terms, though one might add that there is evidence that the industry was also a good deal influenced by fluctuations in the rate of townwards migration of labour. At the same time our knowledge of the structure of the building industry makes it difficult not to believe that in the periods of most intensive activity, like 1825, prospects of high rents were not often less to the fore in the minds of builders than the speculative hope of quick sale on a temporarily rising market. Manufacturing industry was not yet normally a field of joint-stock enterprise and

[1] Matthews, op. cit., pp. 108–10, 204–5.

on that account was, like housebuilding, less directly affected by manias than was investment in transport and public utilities. Fluctuations in investment in the cotton industry, for example, are no doubt largely to be explained in terms of the authors' model, though a more explicit recognition of the innovatory element of investment in this as in other industries (such as iron) might call for certain modifications in the model. Yet in this case, too, the mania element cannot be entirely neglected, if only because there is evidence that investment in manufacturing industry, as also in trade, was on a number of occasions (1792, 1825, 1836) stimulated to an important degree by the mushroom growth of banks and the relaxation of conventional standards of caution on the part of existing banks that were a feature of mania periods.

III

Reference has been made to the special problems raised by the behaviour of investment during the war period. Throughout the war the government's expenditure exceeded its revenue by a large though fluctuating margin. Calculations can be made on the lines indicated by Professor Morgan[1] to show which parts of the government's deficit can be reckoned to have had a direct inflationary impact on the British economy (excluding the effects of expenditure by the government abroad in stimulating exports, which can better be taken account of along with other factors affecting exports). The movements in the series which result from such calculations are very substantial, and although Professor Gayer and his collaborators do make regular allusion to the figures of the excess of government expenditure over government revenue, they hardly appear to give the matter sufficient weight. In their view the chief significance of government deficit spending during the war lay in the encouragement given to exports by military expenditure abroad and loans to allies. Under the influence of this encouragement, and also that of private capital export, exports, as they see it, played a larger part relatively to domestic investment in the major cycles of the war than in later major cycles. Yet Professor Morgan's calculations show that the effect of government deficit spending at home must also have been very great, and the movements involved are often larger than those in exports.[2]

Now the interesting thing is that during the war there is, on the one hand, a marked direct correlation between fluctuations in the size of the government's deficit and fluctuations in the yield on consols, and, on the other hand, a marked inverse correlation between these two and the brick index

[1] E. V. Morgan, *The Theory and Practice of Central Banking, 1797–1913* (1943), pp. 32–34.

[2] In any systematic study of the effects of the war on the level of effective demand, account would, of course, have to be taken not merely of fluctuations in the government's deficit but also of fluctuations in the absolute amount of its expenditure.

(with a tendency for the brick index to lag by about a year). The steep decline in brick production in the mid-1790's and in 1811–14 coincided with steep rises in government deficit spending and consol yields, while the years from 1798 to 1810, during which brick production had a strong upward trend, were a period over which the government's revenue increased more rapidly than its expenditure and consol yields tended to fall.

Two related issues are involved. In the first place, in what sense is it legitimate to describe such periods as 1793–6 and 1811–14 as merely minor expansions? They were characterized by low and declining levels of private investment, but also by mountainous increases in the government's deficit. Likewise can we be sure that 1810 really deserved the name of major boom, in view of the curtailment of consumption that must have been involved by the government's belated success in raising taxation to the level where it at last approximately equalled the enormous amount of the government's current account expenditure? Was not what was involved a switch in productive resources from one use to another rather than a variation in the degree of utilization of productive resources?

It may be remarked in parentheses that a similar problem of interpretation arises in other connexions with regard to the war period. Different sectors of the economy often enjoyed very different fortunes at the same time. The magnitude of fluctuations in corn prices brought about tremendous shifts in real income from one class to another, and the prosperous condition of agriculture that stimulated agricultural investment in the years of high corn prices was not prosperity for the country in the sense normally understood. Rises in the prices of imports and re-exports caused by real or anticipated war-time shortages had favourable effects on profits and confidence amongst merchants and hence perhaps in City circles generally, but rising import prices were the reverse of beneficial to the consumer. Prices of different classes of goods, moreover, as Tooke delighted to point out,[1] often moved in opposite directions according to the effect of the war on the conditions under which they were supplied or demanded. With so many magnitudes moving diversely, the concept of 'the general level of business activity' (and hence of peaks or troughs in same) is evidently one to be used with great caution. In addition to the periods mentioned in the last paragraph, the years 1799–1802 and 1807 may be quoted as examples of times when the course of events in different sectors of the economy was more than usually diverse. The task of marking off the dates of peaks and troughs in such cases and of distinguishing between major and minor cycles consists essentially of balancing one thing quantitatively against another—a purpose for which the available evidence is peculiarly ill adapted. For this reason doubts may be entertained

[1] See, for example, *History of Prices*, i. 189–90, 235, 344–5.

whether, if we had figures for money national income—let alone real national income—they would during the war period always conform to the reference peaks and troughs marked by the authors.

The issue just discussed concerned the *nature* of fluctuations during the war; the other related issue concerns the *causes* of fluctuations in investment at that time. It is natural to ask whether the low level of investment in the so-called minor cycles of the war was not at least as much referable to competition for resources brought about by government expenditure as it was to the overhang of excess capacity from the previous major boom. In their treatment of trends the authors give full weight to this,[1] but they seem to regard it as less important from the short-period point of view.[2]

Further study is clearly needed of the whole question of the relation between fiscal policy, monetary policy, and private investment during the war period.[3] But whatever the exact mechanism involved, if it is agreed that the high level of government spending did somehow inhibit private investment, presumably chiefly through lessening the availability of finance, the question arises why a similar check was not present in ordinary commercial booms in peace-time. If, when government deficit spending rose, the result was not an especially violent boom—for the evidence does not suggest that the periods of highest government war deficit were marked by more than moderately high business activity—but instead was, at least in part, a curtailment of private investment, why was it that increases in private investment in peace-time[4] were not likewise halted by shortage of funds before they had brought about a real boom?

[1] The point is, of course, similar to the idea familiar in Professor Rostow's later writings and elsewhere that, particularly in the period 1870–1914, capital export and domestic investment stood in a competitive rather than a complementary relation to each other *as between* different cycles (or longer periods), though not necessarily *within* cycles.

[2] Those who are sceptical about the meaningfulness of trend drawing may be amused to recall that, while the Columbia authors find no upward trend in brick production during the war years, Professor Shannon, in his original article on the brick index ('Bricks—a Trade Index', *Economica*, 1934, pp. 300–18), found a scarcely less rapid upward trend for the war period than for 1815–49. He referred in corroboration to evidence from Mrs. George's *London Life in the Eighteenth Century* (pp. 79–80) to the effect that the Napoleonic Wars after 1802 constituted a departure from the experience of previous wars by being accompanied by vigorous activity in the building trades—the contrast being explained with reference to the restriction of cash payments.

[3] For example, to what extent should the fall in interest rates between 1798 and 1802 be attributed to greater monetary ease brought about by the Bank Restriction, and to what extent should it be attributed to the greater success of the government in balancing its budget in these years? This is important in so far as the contrast between the fall in private investment in 1793–8 and its rise in 1798–1803 was caused by interest rate movements rather than by a normal cyclical fluctuation in the marginal efficiency of capital following upon the boom of 1792.

[4] And also in 1798–1801, when the government deficit was still very large, though tending to decline.

Evidently there can be no simple answer to such a question, but it may be suggested that broadly the reason was that increases in private investment were able to a large degree to create their own finance by breeding an atmosphere of confidence conducive to an expansion of credit media of all sorts. Government demands for credit did not evoke any such response; when the government required to raise particularly large loans from the public or from the Bank of England to pay for the war, business confidence was not thereby enhanced but rather the reverse.

But there was nothing automatic about the availability of funds for private investment, and there is reason to believe that, particularly in the case of investment expenditure carried out during periods when confidence was not at its peak, this expenditure was partly at the expense of consumption, just as during the war government expenditure was partly at the expense of private expenditure even when it was not paid for out of taxation.

A case can be made for this view in connexion with the aftermath of several booms, notably 1792 and 1836; but much the most important case in point is the great railway boom of the 1840's. The Columbia authors accept without question 1845 as the date when the boom of the 1840's reached its peak, but add (p. 318): 'There is little doubt that railway construction continued for a considerable period after the floatation bubble had burst. The general decline induced by a falling off in the export trade in 1845 was thus considerably cushioned.' In the light of the statistics, this is a notable understatement. The amount of money raised and spent by railway companies was more than £20 million higher in 1846 than in 1845, and there was a further rise of £3 million in 1847.[1] To indicate the orders of magnitude, the fall in the value of exports in 1846 amounted to less than £2½ million, followed by a rise of £1 million in 1847. On the face of it, therefore, it is more than a little anomalous to regard 1846 and 1847 as recession years. The undoubted evidence of declining activity in many sectors of the economy in these years can surely be made intelligible only if one accepts Newmarch's view (not mentioned by the authors) that railway building was paid for largely at the expense of consumption on the part of those who had bought railway shares for a small deposit during the mania period and were then obliged to pay up repeated calls.[2, 3]

[1] For the figures (not quoted by the authors) see T. Tooke and W. Newmarch, *History of Prices*, v. 352.

[2] It is probable that the same sort of thing took place on a smaller scale after the canal boom of 1792–3 and the railway boom of 1836. In the former case the difficulty of raising funds was aggravated by the increase in the general level of costs and prices brought about by the war in the interval between the mania period and the completion of the schemes then initiated.

[3] This seems much more plausible than the line adopted by *The Economist* and other contemporaries, and accepted by the Columbia authors, that the finance of railway building by

The arguments of the last few pages have tended to cast some doubt on the validity of the concept of the general level of business activity, especially though not exclusively as applied to the war years, since there is good reason to believe that on certain occasions the movements of different parts of the national income diverged to an important extent. And in this connexion it should also be borne in mind that even in years when the overall level of effective demand did unquestionably fall, the effect on the volume of production was probably often no worse than to cause a slackening in rate of growth, since in many branches of manufacturing industry the downwards inelasticity of supply was very pronounced. This applies particularly to the latter part of the period.[1]

IV

The important place assigned by the authors to exports, both in their narrative and in their analysis, has already been mentioned. Exports were, in their view, one of the two chief sources of fluctuations in the economy (the other being domestic investment). They were, it is claimed, the most important dynamic force in starting off the major expansions, preparing the way there for the subsequent increase in domestic investment; and their fluctuations caused, and indeed to a large extent comprised, the fluctuations designated as minor cycles. The forces governing the short-run behaviour of exports were, the authors point out, numerous and complicated. Of these forces, two are singled out as having the greatest and most general significance. These are, first, changes in the amount of British capital export, and, secondly, fluctuations in foreign demand caused either by inventory movements or by autonomous fluctuations in the incomes of foreign consumers. As the authors point out, exports at this period consisted very largely of textiles, and investment goods like iron accounted for only a small proportion of the total.

On first consideration of these views, there are two doubts that come to mind. In the first place there is an obvious danger of the historian of this period being biased in the direction of overrating the importance of exports, since the statistical evidence on them is so much more ample than that relating to production for the home market. In the second place the

converting 'floating capital' into 'fixed capital' starved the rest of industry of credit. The theoretical basis for this is extremely dubious, and the comparison of discount rates and calls on railway share capital by months does not support the statement that 'it was the progressive calling up of short-term funds, earlier committed to the railways, that impoverished the money market' (p. 335). The process of railway finance effected a transfer of loanable funds from the country banks to the London money market rather than a reduction in the total supply of loanable funds. Cf. E. V. Morgan, 'Railway Investment, Bank of England Policy and Interest Rates, 1844–8', *Economic Journal* (*Economic History*), 1940, pp. 331–6; idem, *Theory and Practice of Central Banking, 1797–1913*, pp. 152–5; C. N. Ward-Perkins, 'The Commercial Crisis of 1847', *Oxford Economic Papers*, Jan. 1950, pp. 84–87.

[1] See below, p. 23.

authors' neglect of imports both in their theoretical statement and in their treatment of particular periods must be considered a distinct fault. The consequences of large imports in bringing about drains on the Bank of England's bullion reserve and so raising interest rates are ably discussed, but it is nowhere made clear that movements in imports have a direct effect on income levels through the foreign trade multiplier exactly analogous to those caused by movements in exports. It will not do to say without qualification that a rise in exports was responsible for a given improvement in business activity, if we find, as we often do, that it was accompanied by a not smaller increase in the value of imports; in such circumstances the increase in exports can, on the face of it, at least not have been the only inflationary factor at work.[1]

For this reason it will be well in our ensuing discussion of the part played by foreign trade to take both sides of the balance in turn, beginning with imports.

The authors' conclusions about foreign trade are heavily dependent on the average cyclical behaviour of exports, imports, and their different components, as revealed by the National Bureau measures given in Part II of Vol. II. The technique of measurement is the standard one devised by the National Bureau, with certain simplifications and omissions called for by the annual rather than monthly or quarterly basis of most of the series analysed.[2] The standard National Bureau technique divides the cycle into nine stages from trough to trough; with annual data only five are used, these being designated, however, I, III, V, VII, and IX.[3] Stage IX of one cycle is identical with stage I of the subsequent one.

Imports

The behaviour of imports does not at first sight present any features of great note. The following are the figures given by the authors of the aver-

[1] This statement is subject to qualification in so far as the recorded movement in imports was attributable to fluctuations in the level of stocks rather than to fluctuations in the amount consumed. This point is discussed at length below.

[2] The technique may be briefly summarized for the benefit of readers unfamiliar with it. Reference dates are first established for peaks and troughs in general business activity. Each series is broken up into segments on the basis of these reference dates, and the seasonally adjusted data within each segment are expressed in terms of 'cycle-relatives', i.e. as percentages of their average value over the given reference-cycle. Each reference-cycle is then divided into nine stages, and an average is computed of the cycle-relatives in the different reference-cycles at each of these stages. The result gives the reference-cycle pattern. Calculation is also made of 'specific-cycle' measures on exactly similar lines, the time unit here being, instead of reference-cycles, cycles in the individual series considered in isolation. The measures of average behaviour over reference-cycles and specific-cycles are accompanied by measures of the average deviation from the average at each of the nine stages, and by measures of conformity, which relate to the proportion of all reference-cycles to which the series in question is found to respond.

[3] In practice there are often only four really distinct stages, since in the numerous one-year contraction phases figures for stage VII are obtained by interpolation.

age monthly rates of change of the value and volume of imports (reckoned as a percentage of the cycle average) between the different stages of the cycle.[1] (Figures in parentheses represent the average deviation from the average.)

	I–III	*III–V*	*V–VII*	*VII–IX*
Volume	+0·7 (0·5)	+0·7 (0·6)	−1·0 (1·0)	−1·1 (1·0)
Value	+0·5 (0·8)	+0·8 (1·3)	−1·7 (1·6)	−1·7 (1·6)

The authors call attention to the increased rate of growth of values between stages III and V as having a bearing on the regular tendency for the balance of payments to become adverse towards the end of the boom, and suggest that, compared with exports (see below), the demand for imports does not seem to have been too much checked by the rise in prices that might be expected to occur in III–V.

Reference to prices raises the first puzzle about imports. The rate of growth of values shown above increases appreciably in III–V, but that of volume does not. Comparison of the figures yields the implication that import prices on the average fell in I–III and rose in III–V. But on turning to the authors' own new index of prices of imported commodities, we find the following figures of average monthly rate of change:[2]

I–III	*III–V*	*V–VII*	*VII–IX*
0·0 (0·4)	−0·1 (0·4)	−0·3 (0·5)	−0·3 (0·6)

So far from import prices turning up in III–V after a decline in I–III, they remain constant in I–III and actually fall slightly in III–V.[3] It is not clear what conclusion is to be drawn from this surprising inconsistency (not commented upon by the authors) beyond the obvious one that National Bureau average measures of cyclical behaviour are slippery things to handle,

[1] The figures for import values used by the authors are Schlote's, since their work was completed before the publication of Imlah's more elaborate calculations on the subject. Schlote's and Imlah's series on the whole behave very similarly, though occasional differences of substance are to be found.

[2] The figures quoted relate to the movement of the index in its annual form. The monthly rates of movement of the series in its monthly form are as follows:

I–II	*II–III*	*III–IV*	*IV–V*
+0·4 (0·7)	+0·3 (0·6)	−0·1 (0·4)	+0·1 (0·7)

V–VI	*VI–VII*	*VII–VIII*	*VIII–IX*
−0·3 (0·7)	−0·7 (0·9)	−0·7 (0·6)	+0·1 (0·7)

[3] The import price index is not, of course, derived by dividing volume into value, which would make this result impossible, but is an index of prices of individual imported commodities with constant weights.

especially when the deviations from the average are large. The presence of some fault in Schlote's estimates of values may perhaps be indicated. It seems probable that one important source of the trouble lies in the omission of corn from the commodities represented in the import price index, corn being treated as a domestic commodity in the construction of the price indices. Both the price of wheat and the quantity of wheat imports on the average rose steeply in III–V. The average monthly rates of movement of the quantity of wheat imports over business cycles are as follows:

I–III	*III–V*	*V–VII*	*VII–IX*
−0·2 (3·9)	+2·6 (4·9)	−1·4 (6·9)	−2·5 (6·7)

Since wheat imports, on account of their dependent on harvests, must be regarded as on a different footing from the generality of imports, one is led on by these reflections to wonder how the average measures of cyclical behaviour of imports would be affected if corn imports were subtracted from the total. This calculation is not performed by the authors.[1] It is evident, however, that in such a calculation the acceleration in the rate of growth of import values in III–V would be reduced and might disappear, and that the rate of growth of volume might be found to show some slight deceleration in III–V. In any case it becomes clear that one must not think of acclerated growth in value in III–V as the *typical* behaviour of imported commodities.

The authors' explanation of the cyclical behaviour of wheat imports—to digress for a moment to this important subject—is a sensible one. As they point out, the large deviations from the average in the figures quoted above show that no sweeping general conclusions can be drawn.[2] In so far

[1] The statistics available to the authors would not have enabled this calculation to be made for values, but it could have been done for volumes. In this connexion a complaint is in order about the way in which statistics of corn imports are presented in the narrative section of the book. In each chapter of this section Schlote's index figures for the volume of imports for the years involved are tabulated (see above, p. 3, n. 2), and a footnote is appended to the table giving the quantity in each year of wheat and wheat-flour imported, measured in physical units (quarters). Because of the difference in units, it is difficult to assess the importance of the contribution made by corn to the movement of the total. It would surely have been more useful either to have given alongside the Official values of total imports the Official values of corn imports (which are available in the annual *Finance Accounts* in the Parliamentary Papers), or else, if Schlote's figures were to be given for the total volume, to have accompanied them with his series for the volume of grain imports, for the years which this series covers.

[2] In the case, moreover, of a series subject to such enormous erratic fluctuations as corn imports, there is a certain danger of being misled by the method of averaging reference-cycle relatives and the practice of plotting all time-series on a logarithmic scale, which is its counterpart. For example, in such years as 1821–4 the absolute level of corn imports was little above zero, and the fluctuations occurring at this low level had no economic significance; but proportionally, and relatively to the low average for that reference-cycle, these

as the movement is significant, one should probably think of the direction of causation proceeding from corn imports to the state of trade rather than the other way round. That is to say, the tendency of corn imports to fall in VII–III may be taken as some indication that recovery in business was fostered by a preceding period of good harvests—as the authors point out, there was only one major cycle (1816–19) during the first stage of which wheat imports increased substantially. And on the occasions when the later stages of the expansion were marked by high wheat imports, these imports may have contributed to the failure of the boom to last much longer.[1]

The conclusions so far reached about the behaviour of imports do not contain anything very striking. Both value and volume fluctuate in line with the cycle, and it seems that, if corn is abstracted from, there is no clear quickening or slackening of growth as the expansion proceeds. A comparison of the movements of imports in major and minor cycles, however, yields rather more interesting results. The following figures show the monthly rates of change of the volume of imports[2] between the different stages of major and minor cycles respectively:

	I–III	*III–V*	*V–VII*	*VII–IX*
Major cycles	+0·8 (0·4)	+1·4 (0·8)	−1·6 (1·0)	−1·7 (1·1)
Minor cycles	+0·7 (0·3)	+0·3 (0·3)	−0·8 (0·8)	−0·9 (0·7)

The rate of growth of the volume of imports thus increases in III–V of major cycles and diminishes in III–V of minor cycles.

The explanation of this raises matters similar to those discussed in section II above with respect to 'manias'. At or near the peaks of all major cycles in the period except 1845, there took place speculative or semi-speculative booms in the markets for some or all of the chief imported commodities. Steep rises in prices were followed by very large increases in quantities imported and then by equally steep declines in prices. In

fluctuations were substantial, and they therefore produce an appreciable effect on the average measure for all cycles. There is reason to believe that the authors have themselves occasionally been misled by this; they are inclined, for example, to speak of the increase in corn imports in such years as 1824–5 and 1836–7 as if they had an importance which is belied by the absolute amounts involved.

[1] Alternatively, one could seek to explain the tendency for wheat imports to rise in III–V by arguing (*a*) that the early stages of recovery tended to follow and be accompanied by good harvests and low wheat imports, and (*b*) that periods of high and low wheat imports tended to alternate with some regularity (on this see Matthews, op. cit., pp. 30–42). The joint result of these two conclusions would be that, by the time the upswing of the cycle had proceeded some way, a spell of higher wheat imports would be about due.

[2] The authors do not carry out separate calculations for major and minor cycles in the case of the value of imports.

some cases pure speculation was apparently the chief force at work—imports providing, apart from corn, the chief objects of speculation in the commodity markets; but sometimes the expansion of activity in the earlier stages of the upswing seems to have outstripped the supply of imports in such a way as to create a real need to increase stocks. In these circumstances prices naturally rose, and speculation added fuel to the flames to a greater or less extent on different occasions. The large imports thus stimulated put an end to any real or imagined shortage, and prices inevitably slumped, bringing bankruptcy to the speculators in the process. The volume of imports thereupon declined heavily and fell below the level of current consumption. This reaction in turn was frequently excessive (largely because of the general state of discredit prevailing immediately after the upper turning-point of major cycles) and so prepared the way for a subsequent pronounced increase in imports, even if trade was not yet particularly prosperous. The volume of imports would then remain at a more even level until the next major boom approached its peak.

The element of inventory fluctuations in the behaviour of imports is one which the authors emphasize in many passages. The broad outline of events stands out clearly from the sections on the commodity markets in Tooke's *History of Prices*. More information about the movements of stocks might be gleaned from the statistics of the volume of imports entered for home consumption. The authors do not subject these to methodical study; but observation of the comparative mildness or even contra-cyclical character of the fluctuations in consumption of such commodities as sugar and coffee, and also of raw cotton, strongly suggests that in the peak years imports substantially exceeded consumption and that in the ensuing crisis stocks were a good deal drawn upon.

The relevance of this to the measures of rates of movement of imports over major and minor cycles quoted above is as follows. Normally major and minor cycles occurred alternately, and normally major cycles were accompanied by a speculative import boom and reaction of the sort described and minor cycles were not. Hence normally major cycles started in a period of comparative quiescence in commodity markets and ended in a speculative boom and reaction; whereas minor cycles started in the aftermath of such a boom and reaction but did not themselves culminate in one. The more rapid rise in imports in I–III of minor cycles as compared with III–V is explicable by the need to rebuild stocks from the low levels to which they had fallen in the commodity market slump that followed the major boom. Major expansions did not normally start with any such arrears to make up, and the acceleration in III–V reflected the speculative boom with which they concluded. The absence of any clear pattern of acceleration or deceleration in the average behaviour of import volume

over all cycles now become intelligible as the result of the averaging out of majors and minors.[1]

Exports

The average monthly rates of change of the value and volume of domestic exports between the different stages of the cycle are as follows:

	I–III	*III–V*	*V–VII*	*VII–IX*
Volume	+1·0 (0·5)	+0·4 (0·2)	−0·8 (0·8)	−0·6 (0·9)
Value	+0·8 (0·5)	+0·5 (0·4)	−1·3 (0·9)	−1·3 (0·9)

The broad cyclical movement is very clear, but the variety of factors at work makes the explanation of the behaviour of exports in any particular period or to any particular market inevitably complicated.

Inspection of the chart showing the volume of domestic exports over the period indicates clearly that fluctuations were much smaller and less frequent in the 1830's and 1840's than they had been during the war, with the 1820's intermediate in this respect. The disappearance of arbitrary war disturbances on the demand side is evidently part of the explanation; but in view of the much less marked extent to which the amplitude of fluctuations in export *value* diminishes over the period, supply side factors may be inferred to have played a part as well. With factory production progressively replacing outwork, there was a stronger incentive to entrepreneurs to maintain a steady flow of production, even during periods when prices were unremunerative.[2]

The authors begin their general treatment of the behaviour of exports by arguing that the expansion of world population and real incomes,

[1] The above statement is, of course, cast in very general terms. In practice not all major cycles were accompanied by speculative import booms, and such a boom did occur in one cycle designated as minor (1811–16). Moreover, there is one instance of two major cycles occurring in succession (1816–19 and 1819–25) and one or two instances of two minors occurring in succession (1826–9 and 1829–32, and 1801–3 and 1803–8 if one does not include 1801–3 in the same cyclical unit as 1797–1801). What one is led to expect by the argument in the text is this: that there will be acceleration in the rate of growth of the net volume of imports in III–V if a cycle is not immediately preceded by an import boom-and-reaction but contains one; that there will be deceleration in III–V if it is preceded by one but does not contain one; and that the rate of growth will be roughly constant between I–III and III–V if there is one *neither* before nor during the cycle or else *both* before and during the cycle. (In the latter case, one would, to be precise, expect a rapid rise at the beginning and end of the expansion phase, with a less rapid rise in the middle. When only three stages are distinguished, I, III, and V, this would yield a roughly equal rate of growth in I–III and III–V.) Examination of the cases shows that apart from occasional irregularities caused by corn imports and the events of the war, this hypothesis fits every cycle in the period very well, except 1819–25. For some reason the crisis of 1819 was not followed by an immediate increase in imports by way of reaction, but instead dragged out, with the volume of imports low, for several years.

[2] It may be observed that the amplitude of fluctuations in the volume of imports also diminishes after the war, but not so pronouncedly as is the case with exports. The *frequency* of fluctuations in imports does not diminish.

together with the secular fall in British export prices, created a strong upward trend. They then go on to enumerate the following as the chief factors liable to bring about short-run interruptions in this trend:

1. Military factors and other arbitrary obstacles, obviously of prime importance during the war.
2. Changes in the competitive power of foreign producers, caused by tariffs or otherwise.
3. Stringency in the London money market, making difficult the discount of bills drawn on account of exports.
4. Fluctuations in the level of demand from British consumers for exportable commodities.
5. Fluctuations in inventory holdings of British exports abroad, traceable to time-lags.
6. Autonomous fluctuations in the level of incomes abroad, relevant chiefly to exports to the United States.
7. Fluctuations in the amount of British capital export.

Of these factors the last three are held to be most important in the explanation of regularly recurring cycles. Naturally the different factors interact and shade into one another; for example, military events favourable to exports on certain occasions led to an excessive response on the part of exporters and hence to the accumulation of inventories and a recession.

The factors enumerated no doubt go a long way to explain the short-run movements in exports during the period. The authors do not, however, mention one factor which, as I have argued elsewhere,[1] appears to have been of great importance for the period after 1830. This is the fluctuation in the availability of sterling to overseas countries brought about by fluctuations in the value of British imports. In British trade with such areas as the United States and northern Europe, short-period tendencies to deficits or surpluses in the balance of payments could up to a point be met by accommodating capital movements or by shipments of specie. But in the case of overseas areas where financial facilities were not so well developed, sales of British goods appear to have been more closely governed by the availability of bills on London earned by sales in Britain of goods exported from the areas in question. Hence as the value of British imports fluctuated over the cycle, the value of British exports to such areas tended to fluctuate also. This factor helps to explain the regular covariation of export values and the state of trade in Britain. This would otherwise be difficult to explain in the case of exports to remote areas in years when

[1] See Matthews, op. cit., pp. 75–83, where the question is discussed at greater length.

the chief causes of fluctuation in activity in Britain lay in domestic investment.

In so far as fluctuations in exports were due to this cause, the level of exports is to be regarded as a dependent variable, determined by the state of activity in Britain. But the relationship was not, of course, rigid or clear-cut. In some cases fluctuations in exports, even to entirely underdeveloped countries, undoubtedly involved independent inventory fluctuations in the foreign markets (factor 5 in the list on p. 24 above). But fluctuations in this class of exports can to a large extent be explained even on the assumption that all the goods exported passed more or less smoothly into consumption, if regard is had to changes in the availability of sterling.[1]

On the analogy of imports, one might expect to find exports increasing their rate of growth in stages III–V of major cycles and diminishing it in stages III–V of minor cycles, especially if it is accepted that the value of imports had a direct bearing on that of exports. In fact, however, although the pattern is as expected in minor cycles, major cycles, contrary to expectation, also show deceleration in both value and volume in III–V. The following are the average monthly rates of change:

	I–III	*III–V*	*V–VII*	*VII–IX*
Volume of exports:				
Major cycles . .	+0·9 (0·9)	+0·2 (0·3)	−1·4 (1·1)	−1·4 (1·1)
Minor cycles . .	+1·0 (0·5)	+0·4 (0·2)	−0·2 (0·5)	−0·1 (0·4)
Value of exports:				
Major cycles . .	+0·7 (1·0)	+0·4 (0·5)	−1·6 (0·8)	−1·8 (0·7)
Minor cycles . .	+0·9 (0·5)	+0·2 (0·4)	−0·6 (0·6)	−0·6 (0·6)

The deceleration in volume is not especially surprising, if one supposes that either demand or supply or both were fairly inelastic; but the deceleration in value is rather contrary to what one would expect. This result, however, is an average of very different patterns in individual cycles —note the large average deviations shown in parentheses in the table— and its explanation probably lies in the greater variety and complexity of the factors affecting exports as compared with those affecting imports.[2]

[1] The significance of the factor under discussion was undoubtedly much less during the war than it became later. During the war British Government expenditure abroad constituted an enormous erratic element in the supply of sterling. Moreover, the chief fluctuations in exports at that time were to the Continent of Europe, where, as stated above, the availability of sterling was always less important in determining the level of demand for British goods than it was in parts of the world where financial facilities were less advanced.

[2] The only two major booms in which the deceleration of export values in III–V is unequivocal are 1808–10 and 1842–5. The decline in rate of growth in 1810 was due to Napoleon's tightening of the blockade. In 1845 the retardation was due primarily to a minor recession in the United States, and secondarily to a reaction from the boom in exports to the Far East which had followed the conclusion of the Opium War.

From this review of the cyclical behaviour of imports and exports, what conclusions are to be drawn about the role of foreign trade in the mechanism of the cycle?

Movements in imports and exports at the upper turning-points of peace-time major cycles seem to follow a fairly clear pattern. As the boom moved towards its peak there usually came to be felt a real or imagined shortage of some or most imported commodities, and this led to a speculative rise in their prices.[1] Commodity speculation thus accompanied the speculation in joint-stock company promotion which was a regular feature of mania periods. The rise in prices presently stimulated a large increase in the volume of imports, which then outran consumption and caused stocks to accumulate. Exports, in the meanwhile, were stimulated by the increased availability of sterling brought about by the rise in import values, and further stimuli to exports were often present at the same time from investment in foreign securities—a feature of several mania periods—from over-confidence on the part of exporters bred by general boom conditions, and from such other sources as autonomous or partially autonomous booms in the United States.

The rise in the value of exports was not normally so general or on so large a scale as the speculatively induced rise in imports, and a drain of gold from the Bank of England and rises in discount rates were the result. On the other hand, since the increase in imports largely represented stock accumulation and so did not necessarily involve any deflexion of consumers' expenditure away from home-produced goods, the combined effect on incomes of the movements in imports and exports was quite likely to be strongly favourable, even though the balance of payments was moving adversely. Moreover, so long as the rise in import prices continued, it bred confidence in mercantile circles, notwithstanding that its direct impact on the level of effective demand was probably unfavourable.

It was not long, however, before the combined effect of increasing supplies and dearer money caused import prices to fall. This might occur while exports, as well as investment, were still rising, and in this way it served to hasten the end of the period of mania. The fall in prices formed part of the process of pricking the bubble, and had its counterpart in the more or less contemporaneous collapse of joint-stock company speculation. Both were reflected in an early rise in bankruptcies, which frequently

[1] This statement, as a general description of what happened at major peaks, need not be regarded as inconsistent with the equivocal behaviour of the index of import prices between stages III and V (see above, p. 19 and n. 2 there), since (*a*) the latter reflects minor as well as major cycles, and (*b*) the rise in import prices might be of quite short duration, either anticipating the general peak (as in 1809 and 1814) and being followed by a fall in stage V, or coming only at the very end of the upswing following upon a fall or absence of rise in stage IV (as in 1824–5).

preceded the turning-point in business as a whole.[1] These bankruptcies and the associated discredit added an internal drain of gold from the Bank of England to the external drain from which it had for some time been suffering, and there resulted a general crisis of greater or less magnitude. The ensuing fall in the volume and value of imports itself led directly to a reduction in exports, and exports were also likely to be reduced in the course of the reaction from the other abnormal circumstances that had been responsible for their previous rise. The movement in the balance of payments in this phase of the cycle was usually favourable; but as this was achieved largely by drawing on stocks of imports, the net direct effect on incomes was most likely to be adverse. The fall in exports and in the prices of imports formed a large part of what was earlier called the secondary contraction process.

The subsequent recovery in import and export values was also an important element in the expansion phase of the ensuing minor cycle.[2] This expansion consisted essentially of two related elements: a general revival of confidence, accompanied by a fall in discount rates from their crisis levels, and a rebuilding of inventories. Exports were a branch of production particularly affected, (*a*) because exports were affected by import values, and the import trade was specially subject to large inventory fluctuations, and (*b*) because exports were subject to inventory fluctuations on their own account, particularly in the earlier part of the period. In the absence of evidence about the movements of inventories of goods for sale in the home market, however, it would be wrong to assume that restocking in that sphere did not also play an important part in the expansion phases of minor cycles.

This account of the typical minor recovery phase does not provide any explanation for the subsequent contraction. As the Columbia authors rightly point out, the various minor contractions recorded do not in fact seem all to be explicable in the same terms. The minor contraction might represent the delayed primary contraction process of the preceding major cycle (see above, p. 12); it might result from the continuation and reversal of the previous inventory movements in exports or imports; or it might be due to such accidents as a bad harvest. If none of these factors

[1] This explanation of the tendency for bankruptcies to rise in advance of the general turning-point—an interesting fact to which the authors call attention—is no more than a suggestion. Analysis of the bankruptcy returns by trades and localities might cast further light on this, and would be a useful piece of research.

[2] It is interesting to note that in the aftermath of the one major boom that was not accompanied by any speculative rises in import prices, namely 1845, there was after the initial recession no immediate recovery sufficiently pronounced to justify by the authors' criteria the marking off of a minor cycle in the annual data. The relatively moderate rise in the number of bankruptcies in 1846, compared with such years as 1826 and 1837, is also to be noted in this connexion.

was present, the minor recovery might not be followed by a recession at all, but might instead pass straight into the next major upswing (as in 1819–21).

Naturally every minor cycle had its own peculiarities which caused it to depart to a greater or less extent from the pattern which may be thought of as typical. Moreover, not all the fluctuations designated as minor cycles by Professor Gayer and his associates are necessarily to be interpreted along the lines described. For example, the minor cycle 1829–32 lends itself much more readily to explanation in terms of harvest vagaries than in terms of fluctuations in confidence or in inventories, and the minor cycles of the war period were influenced in varying degrees by the special circumstances affecting foreign trade at that time.

It remains now to discuss the part played by foreign trade in the initiation of major expansions. It will be recalled that the Columbia authors, basing their case on the rapid increase in exports in stages I–III of major cycles as compared with the small increases in domestic investment until stage V, regard the increase in exports as the chief stimulant to major expansions in their early stages. This hypothesis is clearly an important one. It is capable of two rather different interpretations, both of which appear to be present in the authors' minds to some degree. According to the first interpretation, the long-run upward trend of demand (referred to earlier in the discussion of investment) was strongest in the field of exports, and it was this trend which brought the economy into a position from which expansion of investment could follow. According to this view, the growth of overseas outlets for the produce of British industry appears as of critical significance for the whole process of the economy's development. The second interpretation does not involve such far-reaching implications as this. According to it, at the opening of major expansions, exports, on account of inventory fluctuations, were as a rule rising at an abnormally rapid rate and so provided a shock which, in the circumstances then prevailing, was sufficient to start a cumulative expansion. This would imply that the origins of major upswings were similar to those of minor upswings.

In order to test the hypothesis conclusively—whichever way it is interpreted—a great deal of information would be necessary which we do not possess. One would have ideally to compare the movements in exports with those in the value of investment, taking account as well of any changes in the propensity to consume and the propensity to import, and one would also have to distinguish between the changes in these magnitudes that were autonomous and those that were induced by fluctuations in income. In practice, of course, this cannot be done, and we do not even have figures

for the money value of investment. Instead a rough-and-ready test may be provided by considering movements in the balance of those items in the overall balance of payments that have a direct inflationary or deflationary effect on income. If this balance does not move favourably in the early stages of major expansions, it follows that the expansion can at least not have been caused exclusively by increases in exports. The balance of trade is not precisely the relevant magnitude to consider, partly because account must be taken of certain of the invisible items in the balance of payments, and partly because fluctuations in recorded imports reflect movements of stocks as well as of consumers' expenditure. For the same reasons comparison of the average rates of growth of imports and exports is inconclusive. Professor Imlah has recently provided us with estimates of the invisible items,[1] and by using these and the available indications of the direction and order of magnitude of inventory movements in imports, it is possible to get some idea of what we want for the post-war years. Let us consider the opening years of each of the major post-war cycles briefly in turn, beginning at the end of the period.

1. In 1843 the net effect of movements in foreign trade on income through the foreign trade multiplier was unquestionably very favourable. Of this effect, more than half was attributable to a reduction in the rate of consumers' expenditure on imported corn. There does appear, however, to have been a net favourable movement in the other items in the balance of payments, and this can best be explained with reference to fluctuations in export inventories, along the lines suggested by the authors. In 1844 the net effect of movements in foreign trade on the level of income was apparently more or less neutral.

2. In 1833 the net effect was favourable, but more than all of the improvement was accounted for by reduction in consumers' expenditure on foreign corn. In 1834 the net effect was slightly adverse.

3. In 1820 and 1821 the value of imports declined while that of exports remained roughly constant. Professor Imlah's calculations accordingly show an improvement in the balance of payments on current account in both years. It appears, however, that in both years appreciable disinvestment in import stocks was taking place, and in the absence of quantitative information on this it is not clear whether the net effect of the movements on the two sides of the balance was favourable or unfavourable.

4. The year 1817 presents a case exactly opposite to the last one. The increase in the value of exports was negligible and that in the value of imports was large; but much of the latter represented rebuilding of stocks after the depression of 1816. The result is again unclear, but is on the face

[1] A. H. Imlah, 'British Balance of Payments and Export of Capital, 1816–1913', *Economic History Review*, vol. v, No. 2, 1952, pp. 208–39.

of it unfavourable to the hypothesis that movements in the sphere of foreign trade initiated the recovery.

Professor Imlah's balance-of-payments estimates have not been carried back earlier than 1816. Comparison of the figures for import and export values, however, gives a picture for 1809 similar to that for 1817, with the rise in imports outstripping that in exports but containing probably a substantial element of stock accumulation. The picture for the upswing 1797–1801 is confused, with apparently an adverse net effect in 1798 but a small favourable one in 1799. It is not till we go right back to the first major upswing of the whole period, that of 1788–92, that we really find unequivocal support for the thesis under discussion. For these years there are only volume figures to go on, but there certainly is an impressive expansion of exports compared with imports.

From a review of these cases it is plain that, since the deflationary effect of increases in imports (especially if corn imports are abstracted from) was so often apparently equal to or greater than the inflationary stimulus received from increases in exports, a substantial contribution—to put it no more strongly—to the initiation of major expansions must normally have been made by factors other than exports.[1] In this connexion the

[1] To guard against misinterpretation, it may be useful to state in more explicit theoretical terms what may and what may not be inferred from the movements discussed in the text. Suppose that there take place equal rises in imports and exports (abstracting from inventory movements), accompanied by some rise in the national income. It follows, as stated in the text, that the rise in income cannot be entirely attributed to the rise in exports, except in the limiting case where the marginal propensity to save is zero. It does not, of course, follow that the rise in exports is not partly responsible for the rise in income. It is even possible under certain conditions that the rise in exports is the chief cause of the rise in income—that is to say, that it is greater than the rise in investment.

The matter may be analysed in terms of foreign trade multiplier theory thus. We write Y for income, X for exports, M for imports, I for investment, m for the marginal propensity to import, and s for the marginal propensity to save. What we want to consider is what may be inferred about the relative magnitude of the change in exports and the change in investment, in the case where there is an equal rise in imports and exports.

If there are no autonomous changes in M, we have $\Delta Y = \frac{\Delta M}{m}$ and also $\Delta Y = \frac{\Delta X + \Delta I}{s+m}$. Since we are considering the case where $\Delta M = \Delta X$, it follows that $\frac{\Delta X}{m} = \frac{\Delta X + \Delta I}{s+m}$. This reduces to $\frac{\Delta X}{\Delta I} = \frac{m}{s}$. That is to say, the rise in exports may be inferred to exceed the rise in investment if the marginal propensity to import exceeds the marginal propensity to save.

If there is an autonomous increase of amount k in imports, we may write $\Delta M = k + m\Delta Y$. The basic equations then become $\Delta Y = \frac{\Delta M}{m} - \frac{k}{m}$ and $\Delta Y = \frac{\Delta X + \Delta I - k}{s+m}$. Where $\Delta M = \Delta X$, it follows that $\frac{\Delta X - k}{\Delta I} = \frac{m}{s}$. This means that the increase in exports is *pro tanto* more likely to exceed the increase in investment if the increase in imports is partly autonomous. If, on the other hand, k is negative so that the increase in the total of M occurs notwithstanding some autonomous fall in M (e.g. lower corn imports following upon a harvest improvement), it is *pro tanto* less likely that the increase in exports exceeds the increase in

importance of good harvests in reducing expenditure on imports has stood out at several points already; and there are many categories of investment, especially but not exclusively investment in stocks, which are not reflected in the available statistical evidence and which may have started rising at an early stage in major expansions.

How, then, should we sum up the part played by exports in the cyclical process? There is no doubt that the value of exports was subject to wide fluctuations. But it is not proved that fluctuations in exports were wider than those in the value of goods consumed at home,[1] nor is it proved that it was exports which controlled the state of business activity as a whole in the periods when the recorded evidence on investment does not show much fluctuation. If it is accepted that short-run movements in exports were partly dependent on those in imports, the way is open for a view of exports which would regard their movements as in some cases more a symptom than a cause of general prosperity, at least after 1815. On some occasions export values did no doubt fluctuate a good deal more widely than other components of the national income. Such occasions occurred not merely during the special circumstances of the war but also in peacetime under the influence of such causes as capital exports, the state of trade in the United States, and inventory fluctuations both in exports themselves and in imports. On other occasions, however, exports and the value of home consumption most probably marched roughly in step, as when both expanded under the influence of general prosperity springing from an investment boom. On yet other occasions, as, for example, when there were particularly good or particularly bad harvests, home consumption was liable to fluctuate violently while exports remained comparatively stable.

A word may be said in conclusion about the contribution made by exports to the upward trend of the economy over the period (see above, p. 28). The rate of increase in the value of exports between 1790 and 1850 was comparatively moderate (0·9 per cent. per annum). Even in the case of cotton textiles, the rate of growth of the volume of exports after 1797 was only very slightly more rapid than that of total production. If any

investment (and, *a fortiori*, that it exceeds the sum of the increase in investment and the autonomous fall in imports, the other two factors which contribute to raising income).

It is obvious that if the increase in exports is itself partly induced—e.g. induced by the increased British demand for imports that occurs during the boom, as suggested on p. 24 above—it is *pro tanto* less likely that the autonomous element in the increase in exports exceeds the increase in investment.

Similar reasoning can be applied to cases where the rises in imports and exports are unequal. Plainly if the rise in imports exceeds the rise in exports, it is less likely that the rise in exports is the main cause of the increase in income, and conversely.

[1] Compare the estimates of sales of cotton textiles abroad and in the home market between 1826 and 1842 in Matthews, op. cit., p. 128.

one factor is to be singled out as having made the greatest contribution to the upward trend in the value of the national income over the period, it should surely be technical progress rather than the opening up of foreign markets. Whether and how and to what extent population growth is to be regarded as an independent factor itself creating the expansion of demand necessary for its own absorption remains an obscure and difficult question always in the background.

ST. JOHN'S COLLEGE,
CAMBRIDGE.

NOTES ON EXCESS CAPACITY

By M. E. PAUL

MR. HARROD has recently argued[1] that the belief that imperfect competition tends to give rise to excess capacity is totally mistaken. This note examines his arguments and finds them unconvincing.

Mr. Harrod's own account of the behaviour of a firm in imperfect competition with free entry is as follows: The entrepreneur 'will plan to charge a price yielding only a normal profit, save to the extent that he is aware of possessing an advantage peculiar to himself, will plan to have equipment on a scale that gives the lowest cost for producing what he can sell at such a price, and, having acquired the equipment, will sell at that price, even although the short-period marginal revenue yielded by such a policy is less than the marginal cost'.[2] A diagram may help us to compare this solution with one along the lines adopted by Professor Chamberlin for the large group.

In Fig. 1 *LAC* represents the long-run average cost curve of the firm, and (1) and (2) are two short-run average cost curves for plants of different sizes. They are plants such that *P*, the lowest point on (1), is also the lowest point of *LAC*, while *Q*, the point of tangency of (2) to *LAC*, is also the point of tangency of the demand curve *d'd'* which will face the entrepreneur when abnormal profits due to initial production at *R* by all members of the group have been eliminated by new-comers. *Q* thus corresponds to Professor Chamberlin's large-group equilibrium point. Mr. Harrod's solution seems to imply that an entrepreneur originally faced with the demand curve *dd* will choose to produce at *P* using plant (1), rather than temporarily at *R*, to be followed by equilibrium at *Q*. Since all entrepreneurs in the group are supposed to choose *P*, there will be no abnormal profits, no new firms will come in, and excess capacity will not arise.

On the face of it, there seems no reason why an entrepreneur should choose *P* rather than *Q* as a long-run position, since in either he is making normal profits; he would surely therefore prefer to make what profits he can at *R*, and have these gradually reduced to normal at *Q*, rather than forgo high profits altogether by starting off at *P*.

Mr. Harrod gives two main reasons for holding that the business-man in these circumstances, when faced with the demand curve *dd*, will choose to start and remain at *P*. The first is an appeal to uncertainty. 'By all accounts and on all hypotheses, the future is largely uncertain. No firm,

[1] R. F. Harrod, 'Theory of Imperfect Competition Revised', in *Economic Essays*, 1952.
[2] Harrod, op. cit., p. 151.

which is interested in a certain line of production, wishes to sacrifice markets available to it for the sake of a fleeting surplus profit. Such a sacrifice will tend to make it weaker in facing the various contingencies of an unforeseeable future.'[1] It is not clear whether Mr. Harrod here wishes to assert that no firm would ever give up future markets, however small, for present profits, however large. He describes the giving up of

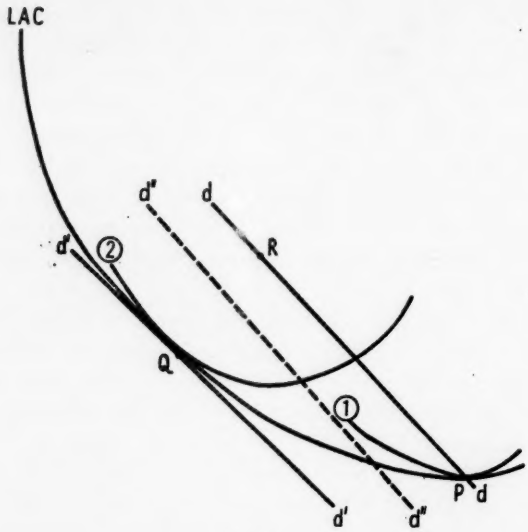


FIG. 1.

future markets as a 'sacrifice', and the surplus profit as 'fleeting', thus turning our attention to cases in which it would be unwise to give up future markets for present profits. Yet surely there are other cases. If the profits were high and not too fleeting, and the loss of market were small, it would clearly be worth going for the profits. It is also not clear what is meant by 'weaker'. It suggests 'more likely to be driven out of business'. Yet a firm in large-group equilibrium need, in this sense, be no weaker than a firm producing at minimum average cost. If stocking manufacturers each produce 500 pairs at 10*s*. a pair, and make 10 per cent. on their capital, why should they be weaker than if there were slightly fewer of them each producing 600 pairs at a slightly lower price and still making 10 per cent.?

Perhaps the uncertainty which is alleged to be inseparable from his choice of *R* is uncertainty about how many firms will be attracted by the

[1] Harrod, op. cit., p. 147.

high profits. Too many may come in, and then some will be driven out before things settle down, and perhaps he amongst them. Unfortunately, however, by choosing P he will not avoid this sort of risk; if others nevertheless choose R, his demand curve also may be pushed leftwards, and despite his caution he may find himself with too large a plant. There seems no reason why uncertainty of the future should lead to a choice of P rather than R.

In his second argument in favour of entrepreneurs choosing to produce at P rather than at R and then Q, Mr. Harrod claims that the accepted adjustment of imperfect competition theory involves an inconsistency between pricing policy and investment policy. Suppose, goes the argument, that the entrepreneur starts from a position in which he can make abnormal profits. He estimates his long-period marginal cost curve, and his short-run demand curve, thereby discovering what is his most profitable output. He installs the plant which will enable him to produce this output most cheaply, charging a price at which short-run marginal cost is equal to short-run marginal revenue; his resulting abnormal profits attract fresh competitors. 'And he will only be able to obtain a normal profit by producing less output than that for which his capacity was designed.'[1] A sensible entrepreneur, however, will anticipate this run of events, and build a plant designed to produce the smaller output which is all that in the long run he can sell. 'This, however, will not save him from inconsistency, if he persists in the maxim of equating short-period marginal cost to marginal revenue; if he does this, he will charge a price that cannot be sustained in the long run without loss of market. . . . This argument may be extended to show that, however small the plant he designs, he will always be pushed backwards, so that it will be redundant. Apparently it is impossible to be an entrepreneur and not suffer from schizophrenia.'[2] It is not, however, clear exactly what is wanted to extend the argument as Mr. Harrod suggests; perhaps once again a diagram will help.

In Fig. 2, B represents the point of optimum profit determined by the intersection of the long-run marginal cost curve and the revenue curve derived from dd, the initial demand curve. (1) is the optimum-sized plant, (2) the best-sized plant for producing at Q, the point of tangency of $d'd'$ to LAC. $d'd'$ is the demand curve which will result from initial production at B. A is the point of maximum profit with demand curve dd and plant curve (2). The entrepreneur who starts at B will then soon regret having built plant (1). However, according to the argument just examined, even if he builds plant (2), should he set his price at A, plant (2) in its turn will become redundant. Similarly, were he to build plant (3), if he set his price at E his plant would again become redundant.

[1] Harrod, op. cit., p. 149.

[2] Harrod, op. cit., p. 149.

This ingenious argument seems, however, to rest on a confusion. It is true that if the entrepreneur produces at *A* with plant (2), and *refuses to lower his price as demand contracts*, he will soon reach a point at which plant (2) is redundant and he could only cover his costs with plant (3). That this may be what Mr. Harrod has in mind is suggested by the phrase quoted above, '. . . he will charge a price that cannot be sustained in the long run without loss of market'. Yet the sustaining of a price which

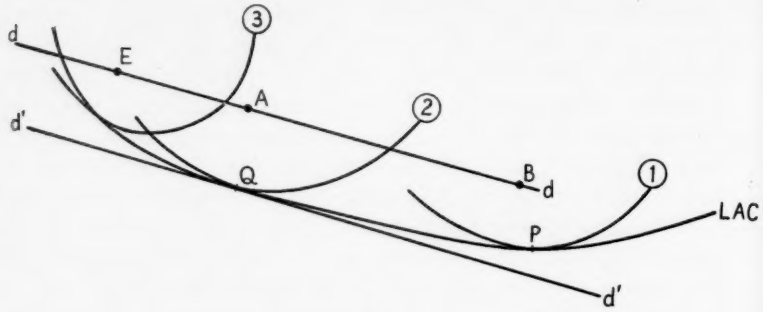


FIG. 2.

once was dictated by the equality of short-run marginal cost and revenue after demand conditions have altered is no part of 'accepted doctrines'. There seems no reason why the entrepreneur who gets what he can at *A* with plant (2) while the going is good should find his plant redundant, provided he lowers price as demand contracts.

We have now discussed the two main arguments by means of which Mr. Harrod seeks to refute the 'accepted doctrine' that firms under imperfect competition tend to be driven into excess capacity. We come now to the claim that these arguments apply not only to firms in oligopolistic imperfect competition but also to firms in imperfect competition of the large-group type, where there is no oligopoly. Mr. Harrod is quite explicit on this point. 'It is to be observed that the foregoing reasoning does not imply that an individual entrepreneur can himself by his own pricing policy affect the appearance or non-appearance of new competitors, i.e. the analysis does not imply oligopoly.'[1] The justification he gives for this claim is as follows: 'If he [the entrepreneur] pursues a prudent pricing policy and his existing competitors do not, his individual action will not prevent the arrival of new competitors, and this may damage him. But what he can do to mitigate the ill effects of the short-sighted policies of his competitors is by charging a low price from the beginning, to make his

[1] Harrod, op. cit., pp. 150–1.

own market less vulnerable. It is of the essence of imperfect competition that the markets of each firm are subject to separate influences.'[1] In terms of Fig. 1 this account of the matter suggests that if one wise entrepreneur produces at P, while the rest start at R, his demand curve will be pushed to the left by new competitors, but not as far as the demand curves of the rest of the group. This, however, by no means implies that choosing P is therefore the wisest policy for an entrepreneur to pursue; if his demand curve is pushed to $d''d''$, beyond the point where he can cover costs with his ambitious plant, he will be caught, and regret not having charged R and built a sub-optimum plant along with the rest. There is thus no reason to think producing at P a prudent policy; for it to be so requires not merely that the market of each firm should be subject to its separate influences, but also that it should be subject to no common influences. If by producing at P the entrepreneur could prevent the appearance of all competitors whose goods rival his particular brand, he would thereby achieve a permanent equilibrium free from some of the uncertainties accompanying an influx of new firms. But this requires that he should have large oligopolistic powers. Once it is admitted that his demand curve may be pushed back even if he starts at P, P becomes as temporary an equilibrium point as R; the task of calculating what size of plant will finally be needed is as precarious as at R, and temporary profits are forgone for nothing.

Nor is it clear that in every case by starting at P the entrepreneur will end up with a demand curve to the right of that which would have resulted had he started at R. The goodwill which he acquires by charging a lower initial price may prevent his immediately losing his share of customers to the new-comers, but as their prices fall to a level nearer his own, and as firms appear whose goods are more nearly substitute to his own, it seems possible that his demand will eventually fall as far from P as it would have done from R.

Once the assumption of oligopoly is removed, it can also be shown that should an entrepreneur succeed in establishing himself in splendid isolation at P, he would not wish to remain there. For at P, marginal cost exceeds marginal revenue, so there is an apparent gain to be made by restricting output and raising price; only fear that this policy would cause a shift in the demand curve by attracting new entrants could deter an entrepreneur from pursuing it. There would thus be nothing to prevent a competitor who is without influence on potential rivals from raising his price. This leads to the curious result that a large-group competitor whose demand curve gets out of line with those of the rest of the group, remaining at dd while theirs shift to $d'd'$ must be either a monopolist or an oligopolist. If

[1] Harrod, op. cit., pp. 150–1.

he can raise price above P and restrict output without attracting newcomers he is a monopolist. If, by thus making abnormal profits he does attract rivals, pushing his demand curve to tangency, he is an oligopolist. Non-oligopolistic imperfect competition does not seem capable of explaining the behaviour of one firm which strays from the fold.

Mr. Harrod says that his views are a development of a certain line of thought put forward by Mr. Kaldor in 1935.[1] Yet on this question of excess capacity in non-oligopolistic situations, what appear to be the most relevant parts of the latter's article constitute a clear defence of accepted doctrines.

Mr. Kaldor develops the now familiar notion of potential competition as a possible preventative of excess capacity in oligopolistic situations. If a producer takes into account the effect of his price, not only upon existing competitors but also upon those who might come into the field if prices and profits are high, he may well act as if his demand curve were much more elastic than in fact it is, and this may lead him to charge a price low enough to avoid being driven into excess capacity. Mr. Kaldor is at pains to show, however, that potential competition cannot affect the issue in the case of a large group in which there are no oligopolistic influences. In such a group it is assumed that consumer's preferences are evenly distributed among the different brands produced by the various members of the group. Thus the results of a change in price by one producer will be evenly spread over all other members of the group, materially affecting none. If preferences are spread in this even fashion, says Mr. Kaldor, potential competition will not enter into the picture. For although all producers as a body could, if they wished, lower their prices sufficiently to prevent new competitors from coming in, 'since the appearance of any single new producer will only affect the demand of a single existing producer very slightly, while similarly the reduction of price of a single existing producer will only slightly affect the profits which a potential producer can expect, no producer could take these indirect effects on his own price policy into consideration'.[2] Mr. Kaldor, however, goes on to assert that even where numbers are large, it is unrealistic to assume anything like an even distribution of consumers' preferences over the whole field. Each producer will have some rivals whose goods are fairly close substitutes to his own, and other less close rivals whose price changes affect him less. 'Savile Row tailors will be most influenced by Savile Row prices; they will be less concerned with fluctuations in the price of East-end clothes.'[3] So even where there are many producers, each one can affect the policy of some by his own price changes, and may therefore influence the arrival of

[1] N. Kaldor, 'Market Imperfection and Excess Capacity', *Economica*, 1935.
[2] N. Kaldor, op. cit., p. 36.
[3] N. Kaldor, op. cit., p. 38.

potential competitors.[1] Mr. Kaldor is thus able to apply potential competition to the case of many producers only by saying that a large group is essentially similar to a small group, by virtue of each firm in both having the power by itself to influence some others. He brings out clearly how potential competition cannot affect the issue in the absence of oligopoly.

It seems of some importance to show that the large non-oligopolistic group at all events is not affected by potential competition. For once oligopoly is admitted all things become uncertain, and potential competition may well fail to eliminate excess capacity. Mr. Kaldor, for example, although proclaiming the ubiquitous character of oligopoly and with it of possibilities of potential competition, by no means denies the tendency towards excess capacity. Potential competition *may* prevent firms being driven into excess capacity, but if producers do not exercise sufficient foresight, are not sufficiently confident of the reactions of others, excess capacity may still arise. '... the effect of the competition of "new entrants" and consequent reduction of the level of profits earned may take the form of a rise in costs rather than a reduction of prices'.[2] In so far as each producer considers his influence on others already in the field, oligopoly will lead to a price higher than in the non-oligopolistic large group; in so far as he considers his influence on potential competitors it will be lower. Thus the final result of potential competition in oligopolistic situations is uncertain. Could the non-oligopolistic large group, however, have been shown to be affected by potential competition, in the absence of the price-raising tendencies of oligopoly, a decisive force against excess capacity might have here emerged.

Although we have so far found no reason to abandon the doctrine that imperfect competition leads to excess capacity, yet the introduction of short-run cost curves clearly makes necessary some modification of the traditional path towards equilibrium. By leaving out short-run curves Professor Chamberlin's solution ignores the fact that a firm cannot experience changes in output large in relation to its long-run cost curve without getting caught with the wrong-sized plant. Even when a firm cannot affect the entry of rivals by its own actions, it may foresee new entry caused by the price policies of the whole group. Granted such foresight, it would do better to build plant (2) (Fig. 2) than plant (1), even though

[1] It is not proposed to discuss here how far Mr. Kaldor is right in his suggestion that non-oligopolistic imperfect competition never occurs. It is assumed that the Chamberlin large group is of theoretical importance, because it may approximate certain areas of reality. In support of this, it may be pointed out that a single Savile Row tailor may well have little idea which of his Bond Street and Regent Street fellows will take notice of any price changes he may make, and in fact might easily assume they will ignore what he does.

[2] N. Kaldor, op. cit. p. 46.

profit maximizing in relation to the long-run cost curve alone would indicate producing at *B*. Introducing the short-run cost curves thus may involve initially a different-sized plant, price, and output from those which would be chosen were the long-run cost curve alone in question. The equilibrium adjustment when abnormal profits have been eliminated is still, however, the accepted one, involving excess capacity.

THE PROCESS OF IMPERFECT COMPETITION

By J. R. HICKS

1. THE purpose of this paper is to discuss certain aspects of the relation between the theory of imperfect competition and the Marshallian doctrine of the short and long periods. Consulting the Scriptures, one finds that this is a matter which hardly arose on the approach adopted by Professor Chamberlin; while Mrs. Robinson, though she was undoubtedly alive to the problem, said nothing about it that was very consecutive. Thus it was left to Mr. Harrod, in his 1934 paper,[1] to give what has long remained the classical analysis. Nearly all economists who have been teaching this subject during the last twenty years have taught what they had learned from that paper; certainly I have done it myself. This prop and stay is now taken from us. For in the new paper[2] which is added to the old one in *Economic Essays*, Mr. Harrod goes a long way in retreat from his former position. It is, I think, impossible to defend the old theory from some of the criticisms which its author now brings against it. It has to be replaced by something different. I do not, however, feel that Mr. Harrod has completed the construction of the new theory which his latest work causes us to envisage. Some essential elements seem still to be missing. It is in the hope that I may be able to provide some of these elements that I have written the present paper.

I propose, quite deliberately, to be rather formal. For what I am looking for is a model which will provide a skeleton for the new theory, just as the familiar cost and demand curves (with their derivatives) did for the old. I shall not be much concerned with the clothing of the skeleton. All I am seeking to do is to construct a model that is usable, and that will start us thinking in sensible directions.

2. Consider the position of a firm which is proposing to set up a new plant. Time, as it stretches forward from the date of decision, will be treated as divisible into three *periods*. The first is the *construction period*, in which output will in any case be zero. The second I shall provisionally call the *close period*; during this period the plant, if built, will be completed and in action, but competitors will not have had time to get similar plants into action. The third, which may similarly be called the open period (both because it is of indefinite length and because the market is now open to competitors), begins at the date when competitors will have had time to get similar plants into action, if circumstances are such that it is to their

[1] 'Doctrines of Imperfect Competition' (*Q.J.E.* 1934); reprinted in *Economic Essays*, p. 111.

[2] 'Theory of Imperfect Competition Revised', *Economic Essays*, p. 139.

interest to do so. For simplicity I shall assume that a single price-and-output policy has to be chosen for the close period and a single price-and-output policy for the open period. In other words, variations in price within the close period or within the open period (though they may of course occur as a result of unexpected changes in the market situation, when the time comes) are not envisaged at the time the plan is drawn up. I hope that the reader will be able to accept these as reasonable simplifications.

I shall (as is the common practice) assume that factors can be hired (or acquired) on the market at constant prices, so that cost functions are determinate. I shall then proceed to identify these cost-functions with the Marshallian cost-functions of the *short* and the *long* period. This is admittedly a drastic simplification; but I think it should be acceptable, at least for many applications if not for all, on the following ground.

It is reasonable to assume that the plant which is being put up will be fairly durable—of a durability which is long in relation to the length of the close period. The technical necessities of production by modern techniques will ordinarily ensure that this is the case. It follows that the size of the plant which must be constructed will be related to the output which is planned for production in the *open* period. The output (per week) in the open period will then play the same part in our theory as Marshall's long-period output, while the output (per week) in the close period, being produced from a plant which is not necessarily in perfect adaptation to it, will be a short-period output in Marshall's sense. I shall accordingly allow myself to use the familiar terms 'short-period output' and 'long-period output' for the outputs per week of the close and open periods respectively. Thus in our theory, as in the old theory, the significant variables may be described as short-period output and long-period output; it is these which determine the production plan of the enterprise.[1] But we must be very careful to distinguish cases in which expectations of future market conditions are formed in one way or another.

3. On this approach, the classical (or Harrod 1934) theory of imperfect competition appears as the particular case in which the market conditions confronting the firm are expected to be the same in the open as in the

[1] It is not meant to be implied by this framework of assumptions that the output of the firm is necessarily homogeneous. A considerable degree of heterogeneity can be allowed for by supposing the figure which represents the *output* of the firm to be an *index* of output, and by assuming that the composition of the output which corresponds to any given value of the index is determined independently of the considerations which are our present concern. This simplification is of course quite out of the question when we are specially concerned with problems of the composition of output; but for present purposes, when we are solely concerned with *scale*, there does not seem to be any harm in it. It is, I think, a great mistake to conclude that the *theory* of the single-product firm is made inapplicable to reality by the empirical fact that most firms produce a variety of different products.

closed period—the simplest case out of a variety of alternatives. If we were to adopt Dr. Lange's terminology,[1] we should call it the case of static expectations; but for reasons which will appear later I should prefer to say that in this case expectations are *simple*. I need not repeat the well-known argument which shows that on this assumption short-period output (determined from the condition that short-period marginal revenue equals short-period marginal cost) and long-period output (determined by the corresponding long-period condition) are equal. I shall indeed be coming back to that argument in a different connexion. Here we may merely note that simple expectations engender a flow of output which is itself stationary.

The theory of market equilibrium which is based upon this type of behaviour by the firm will then run as follows. It is assumed (for there is no advantage in assuming the contrary) that actual demand during the close period coincides with expected demand. Nothing will therefore happen, during the close period, to impede the realization of the plan. During the (realized) close period there does therefore exist that *static* form of excess capacity which is represented by the downward slope of the average cost curve[2]—a downward slope which is established by the tangency between that short-period average cost curve and the long-period average cost curve, so long as the latter is taken to be downward sloping. Such static excess capacity as this is fully compatible with equilibrium.

Before going farther it is necessary to distinguish between the case of the marginal and that of the intramarginal firm. If our firm is a marginal firm, it will not make significantly more than normal profits during its close period, so that there will be no reason why it should have to face additional competition when the market becomes open. Thus there is no reason why the expectation of unchanged demand in the open period should not be verified. The close period situation will maintain itself in the open period. The same output will continue to be produced, and the static excess capacity will remain.

If our firm is an intramarginal firm, its profits (during the close period) will be in excess of normal, and it is therefore to be assumed that the commencement of the open period will be attended by the arrival of competitors. The demand curve for our firm will shift to the left, so that the output which it will produce will be less than that which it had initially planned for the open period. Output is thus still smaller, in relation to the capacity of the plant, than it would otherwise have been. But the excess

[1] O. Lange, *Price Flexibility and Employment*, p. 1.

[2] Average total cost, in Mr. Harrod's terminology. But I cannot refrain from a word of protest against this term, which I have long disliked. Total cost is usually employed to mean average cost multiplied by output; it is a thorough nuisance to have to use it in another sense. Could we not agree to say 'average inclusive cost', which is clear and unambiguous?

capacity which arises from this cause must be sharply distinguished from the other. That was a static, but this is a dynamic phenomenon—it arises from a discrepancy between actual demand conditions and those which were planned for. It is a phenomenon of miscalculation and disequilibrium.

I suppose that the story would then be continued as follows. As time goes on, the industry will 'fill up'; the point will ultimately be reached when no further profit is to be got from entry by a new firm. When this point is first reached the industry will consist of (1) the latest arrivals, which are approximately marginal, (2) the old firms, which were intramarginal. So far as the marginal firms are concerned, they are in equilibrium, with static excess capacity, but no dynamic excess capacity, straight off. So far as the older firms are concerned, they must continue with their dynamic excess capacity until their plant wears out; when it does so, they replace it with new plant, more appropriate to their now reduced circumstances. In effect, when they do so, they convert themselves into marginal firms; they abolish their dynamic excess capacity, but their static excess capacity remains with them. It is in this way that the industry is supposed to settle into its *long-period equilibrium*.

4. This, I think, is the classical story; it is this story which Mr. Harrod now seeks to modify—by abandoning the assumption of simple expectations. It must at once be granted that this assumption is a particularly weak link in the preceding argument. It does not need much imagination to realize that entrepreneurs are much more sensitive to the danger of potential competition—which affects the anticipated long-period demand for their products—than they are to the elasticity of short-period demand. I find myself in hearty agreement over the need for abandoning this particular assumption.

But when we abandon it, what do we put in its place? There is only one sort of simple expectation, but the varieties of complex expectation are many. It is not possible to analyse any of these varieties properly without a formal extension of the existing theory of the firm. For as soon as the assumption of simple expectations is abandoned we lose the power of determining short-period output and long-period output separately; they have to be jointly determined. We need a theory of this joint determination before we can get any farther.

5. As a first step in the construction of that theory we have to face the old riddle: what does the entrepreneur maximize? Corresponding to any choice of the two key variables there will be a particular rate of short-period profit, and a particular rate of long-period profit, that can be expected: how does he weigh one of these against the other? It seems to me that we ought to think of him as maximizing a certain combination of these

two profits. Thus if g is the rate (per week) of short-period profit, and G is the rate (per week) of long-period profit, he maximizes $lg+mG$, where l, m are *weights*. If we remember that price and output are taken to be constant *within* each of the two periods, and apply the ordinary rule of maximizing the present value of the stream of profits, this is in fact the result which we get. (On this derivation, we implicitly assume a perfect capital market; but it is not evident that the assumption of an imperfect capital market would make any difference to the result.)

The weights l, m may then be taken to be governed by such things as the lengths of time for which the close (short) and open (long) periods are expected to last, the rate of time-preference, and the willingness to bear risks. Such subjective factors (and capital-market factors) may well, on occasion, be of great importance. The entrepreneur who is mainly interested in seizing a quick profit (I shall call him the *Snatcher*) will have a high l relatively to his m; the one who is mainly interested in building up a steady business (I shall call him the *Sticker*) will have a high m relatively to his l.[1] Such types of psychology are, I think, readily recognizable in the pages of business history. I shall simplify by treating these weights as permanent characteristics of particular firms, so that they can be reckoned as constants; but it must of course be noticed that to do so is to make a drastic simplification.

I shall continue to use small letters for short-period magnitudes and capitals for long-period magnitudes. Thus x and X will be short-period and long-period output respectively. r and R will be the (total) revenue per week expected from selling these outputs; c and C will be the (total) costs per week, assumed capable of allocation between periods. Thus what the entrepreneur maximizes is

$$lg+mG = l(r-c)+m(R-C).$$

In order that this should be a maximum, its partial derivatives with respect to x and X must be zero. Using the suffix notation for partial derivatives, this gives us two equations

$$l(r_x-c_x)+m(R_x-C_x) = 0$$

$$l(r_X-c_X)+m(R_X-C_X) = 0$$

to determine the two key variables x and X.

This is the complete generalization of the 'marginal revenue marginal cost' equation; but it is usually capable of some simplification. In the first place, since the firm is under no compulsion to disclose its open-period price-policy during the close period, the effect of the planned long-period

[1] I must acknowledge the debt which I owe, for this distinction, to reflection upon Mrs. Paul's paper, appearing elsewhere in this number.

output on short-period revenue can ordinarily be taken to be zero. Thus $r_X = 0$. Further, since we are making the usual Marshallian assumption that the size of plant is determined by long-period output only, short-period output will have no effect upon long-period cost. Thus $C_x = 0$. The equations accordingly reduce to what I shall regard as their standard form

$$l(r_x - c_x) + mR_x = 0$$

$$l(-c_X) + m(R_X - C_X) = 0.$$

The 'straight' parts of these equations (the term in l in the first, and that in m in the second) correspond to the ordinary relations of marginal revenue and marginal cost; the 'cross' parts are the qualifications which need to be inserted in the present more general case. It will be noticed, as a convenient feature of the model with which we are working, that no more than one 'cross marginal revenue' and one 'cross marginal cost' survives in the standard form of the equations, so that these terms can be used for R_x and c_X without ambiguity (though also without symmetry).

These equations have now to be interpreted in the light of the familiar relations (already utilized) between short-period and long-period cost. For the purpose of the present analysis these relations would be presented in the following way. We have seen that c (short-period total cost) is a function of x and X, while C (long-period total cost) is a function of X only. But we know that if X is chosen in such a manner as to make c a minimum for given x, x and X will be equal, while c will then be equal to C. Thus if $x = X$, $c_X = 0$ (since c is minimum) and $c = C$. Now C_X (long-period marginal cost) is the derivative of total cost against a change in X, x being kept equal to X, and c_X being therefore equal to zero throughout. Thus $C_X = (dc/dX)$ (the *total* derivative) when $x = X$. But

$$\frac{dc}{dX} = \frac{\partial c}{\partial X} + \frac{\partial c}{\partial x}\frac{dx}{dX}$$

(by the usual rule for a total derivative). Here, when $x = X$, $(\partial c/\partial X) = 0$, $(dx/dX) = 1$. Thus $C_X = (dc/dX) = (\partial c/\partial x) = c_x$. Thus we have the additional information that when x and X are equal, short-period and long-period marginal cost are equal ($c_x = C_X$); while the cross marginal cost (c_X) is zero.

6. As a first step in the application of this model, we may verify the classical analysis for the case of simple expectations. The characteristics of this case are (1) that R_x is zero, since long-period demand is unaffected by short-period output (or price-policy); (2) that if $x = X$, $R = r$. The cross-term in the first of our standard equations drops out in the light of the first of these conditions; that equation accordingly reduces to the

familiar equation of short-period marginal revenue and short-period marginal cost. The second equation is then satisfied by $x = X$. For if $x = X$, the cross-term in the second equation drops out (as we have just seen), while the remaining terms become identical with their short-period counterparts. $C_X = c_x$ (as just established); $R_X = r_x$ (by the assumption of simple expectations). Thus if $x = X$, and either of the two standard equations is satisfied by the common value, the other will be satisfied also. But so far we have offered no more than an algebraic proof of what was well known already.

We now strike out into less well-charted seas. The next assumption which I should like to examine is that in which the entrepreneur's expectations remain static, so far as the general state of demand, for the products of the industry of which he is a part, is concerned; but in which he expects the arrival of competitors in the open period—competitors who will arrive independently of anything which he can do by his short-period policy. This is not one of the hypotheses to which Mr. Harrod particularly desires to draw our attention (in his 1953 paper). He is (very naturally) anxious to rush on to the discussion of *growth*; and so far as he will admit the case of expectations that are in any way static, he is mainly interested in the possibility that the firm may seek, by its short-period price-policy, to keep out competitors. Now conditions in which the firm is able, by its short-period price-policy, to influence the behaviour of competitors, to keep them out of (or let them into) its own particular market, seem to me to have a distinct flavour of oligopoly about them. I shall therefore call that case the case of *oligopolistic expectations*. The present case, which comes before it in our present procedure of generalizing the classical theory of imperfect competition, I shall call the case of *polypolistic expectations*.

7. With polypolistic expectations, long-period revenue remains unaffected by short-period policy, so that it is still true that the cross marginal revenue is zero. The cross-term in the first equation still drops out. The first equation still reduces to the familiar equation of short-period marginal revenue and short-period marginal cost. But the second equation cannot be relied upon to simplify similarly. It will, however, simplify approximately, if our firm is very much of a Sticker, with m large relatively to l; the cross-term in the second equation will then be small, and the second equation can be represented, to a sufficient approximation, by the equation of long-period marginal revenue and long-period marginal cost. Both R and C are now functions of X only, so that this equation suffices to determine X (and with it the size of plant) directly. But since the firm expects a smaller demand for its product in the open period than in the close period, the size of plant will be smaller than it would have

been apart from the expectation of future competition. Polypolistic expectations make for smaller plants, which draw less of the economies of large scale, than would be put up if expectations were simple.

The size of plant being determined, the first equation (of short-period marginal revenue and short-period marginal cost) remains to determine short-period output. x and X are now so far independent that we can think of the two equations being represented, in the usual manner, on the same diagram. Short-period demand being greater than the expected long-period demand, short-period marginal revenue (for a given output) will be greater than long-period marginal revenue. Short-period output will accordingly be greater than the planned long-period output. Though the firm will still be planning for a situation in which it will have static excess capacity in the long period, it is not inevitable that such excess capacity should exist in the short period. Short-period output may well be in excess of that which would minimize cost from the chosen plant.

If, however, our firm is less of a Sticker, the size of plant which it installs may be influenced by the fact that it does not propose to produce its short-period output from the plant which is best suited for producing that output. The term $l\ (-c_X)$ may then become appreciable. Now if $X < x$, c_X must be negative; an increase in the scale of the plant will lower short-period total cost. It follows that if any attention is paid to this element, it will work in the direction of increasing the scale of plant chosen. It will thus do something to offset the tendencies described in the previous paragraph, but it seems most unlikely that it will go so far as to offset them altogether.

8. What is the theory of market equilibrium which corresponds to this case of polypolistic expectations? Once again it is necessary to distinguish between the marginal and the intramarginal firms; but it is no longer possible to define a marginal firm as one which expects to make no more than a normal profit. For each sort of firm will expect to make a larger profit in its close period than it will be confident of making in its open period. Thus those expectations which will just induce a firm to come in will be for a *more than normal* profit in the close period just offsetting a *less than normal* profit in the open period. On that basis it is just worth while to come in; but it will be more tempting to come in on that basis if one is a Snatcher, being specially interested in short-period profit. Marginal firms are thus very likely to be Snatchers. I think that this conclusion accords very well with experience, which finds a very common characteristic of an industry with genuinely free entry to be a rapid turnover of marginal firms.

This rapid turnover is not necessarily a wasteful process; but I think we are justified in regarding it as being, in general, a source of waste. But the

wastes which are associated with it are not such as can be analysed in terms of static excess capacity. So long as these firms can produce profitably there is no reason why they should not produce more or less up to capacity; once they cease to be able to do so, they are heading for disaster.

So far as the intramarginal (and comparatively stable) firms are concerned, their expectation of future competition will prevent them from over-expanding their plants in the way they were supposed to do in the 'simple' theory. In their close period (immediately after their plant is put up) there is no reason why they should not produce up to capacity; later on, when their competitors have arrived, they will (if they have calculated aright) be producing no more than the output for which their plant was planned, so that the usual excess capacity of the static kind is, formally at least, present. But it should be noticed that all that has happened at this stage is that the dynamic excess capacity of the simple expectations case has been transmuted into a static excess capacity. And this, indeed, is not the end of the story. When such firms come to replace their plants, they will still (so we may suppose) have polypolistic expectations, which will cause them to reduce their plants to that size which would be appropriate to them if they were marginal firms. The 'static' excess capacity accordingly disappears; it turns out to have been 'dynamic' after all!

We must, I think, conclude that the real evils of this sort of industrial organization (and industrial psychology) are very inadequately expressed in terms of static excess capacity. The real wastes, which we should expect to find in practice, are much more obvious, much less highbrow. They consist in the turnover of marginal firms, and in the inability of the steady firms to expand their plants to such a size as will enable them to take full advantage of the economies of large-scale production.

9. I now turn to the case of *oligopolistic expectations*, where the firm considers that the amount of competition which it has to expect in the open period is a function of its own close-period price-policy. This gets us nearer to the field in which Mr. Harrod is interested (and Mrs. Paul also); though it should be noticed that we are still retaining the assumption that the general state of demand, for the products of the industry as a whole, is expected to remain static.

As before, we commence our study by a consideration of the standard equations.

The cross-term in the first equation is now no longer negligible; an increase in short-period output (a reduction in short-period price) may now suffice to deter competitors, and may therefore increase the demand for the firm's product in the open period. The cross marginal revenue will then be positive. This means that the classical equation of short-period marginal revenue to short-period marginal cost is liable to under-estimate the

short-period output which it will be profitable to produce. But there is more to be said on the matter than that.

It is possible, if our firm is very much of a Sticker, that it may be largely uninfluenced by the straight terms in its first equation; its short-period output may be mainly determined by the effect of that output on long-period sales. We can even think of an extreme case (that of the *Perfect Sticker*) in which short-period output will be solely determined by the condition $R_x = 0$.

What would be the practical meaning of the equation in this limiting form? It is not to be supposed that the firm could go on expanding the long-period demand for its product by indefinite increases in short-period output. There will be some short-period price which will be so low that it does all that can be done in the way of deterring competitors; once short-period output is greater than the output corresponding to this critical price, the cross marginal revenue is *necessarily* zero. Thus the condition $R_x = 0$ merely implies that short-period output must be kept greater than this critical output; but it does not tell us how much greater. To that extent short-period output appears to be indeterminate. Yet there is in fact no point in reducing price below the critical price; so that it seems fair to think of the Perfect Sticker as fixing a price which is low enough, in his view, to deter competitors, and then selling what he can sell at that price. Now this is a kind of behaviour which begins to look, in a certain sense, familiar. One begins to ask whether it is not the case that the firms which have been interviewed by Mr. Andrews are firms which are (or think themselves to be) Perfect Stickers?

It is entirely consistent with the present theory that most firms should be Perfect Stickers; whether or not that is so is an empirical question. But theory can nevertheless insist that the Perfect Sticker is a theoretically extreme case; that for anyone who gives any weight at all to the l terms in his equations the costs of the policy we have been describing may be very formidable. Mrs. Paul is surely right in maintaining that there is nothing inconsistent in being a Snatcher and having oligopolistic expectations (for that, I think, is what it comes to); and that an entrepreneur who gives fair weight to both short-period and long-period considerations may well decide, on occasion, that the output needed to deter competitors will be, in the short run, too expensive to produce. He must then resign himself to the probable appearance of competitors; in spite of the chance which he had of acting otherwise, he will in fact act as if his expectations were polypolistic.

10. Let us return to the Perfect Sticker and look at his second equation. He is doing all he can to put off potential competitors, and the long-period demand which he expects must be established on that basis. Since he

neglects l terms, he will determine the size of his plant solely on the basis of long-period marginal revenue (which is now a function of long-period output only) and long-period marginal cost. His equations have accordingly reduced to the limiting form

$$R_x = 0, \qquad R_X = C_X.$$

The behaviour which is represented by these equations is worth a little consideration.

It may conceivably happen that the firm believes in its ability, by keeping price sufficiently low, to keep out competitors altogether. If so, it will follow that the choice of a short-period output which is consistent with the first equation will cause the long-period demand curve confronting the firm to be the same as the short-period demand curve. That is to say, though its expectations would not have been simple if it had acted differently, it has got itself into a position which is such that it can allow itself to have simple expectations. The complete position which it will take up will therefore be the same as that which it would have taken up if its expectations had been simple. It will plan for an equal output in the close period and in the open period; in consequence, *though it does not know that it is doing it* (*or anything equivalent*) it will in fact act in such a way as to make short-period marginal revenue equal to short-period marginal cost. Thus, provided that short-period output is greater than the critical output, it does not after all matter whether the firm is a Sticker or a Snatcher; it will act the same way in either case. And the way in which it will act is precisely that which emerged from our study of simple expectations. If the firm were a marginal firm, it would then follow that it was in a state of static excess capacity; but a firm which was in the situation we have just been describing does not look much like a marginal firm—its position is hardly distinguishable from that of the textbook monopolist!

The somewhat paradoxical conclusions which we have just reached are indeed to be attributed to the fact that we have been examining an extreme case. It is very extreme to assume that our firm can get into such a position, by its low price policy, that it can be *sure* of excluding competitors. The most we should expect, in an industry which in any way deserved the name of competitive, is that it would feel that it had a good chance of keeping them out. If so, it could not attain to the position we have been supposing, in which it could be treated as having *simple* expectations; the most which could be expected would be that its expectations would be mildly polypolistic. It would then work on the assumption that long-period demand would be somewhat less than short-period demand, but perhaps not so very much less. It would then follow (applying our analysis of the polypolistic case) that its short-period output would be

somewhat in excess of its planned long-period output; and that excess might be quite sufficient to abolish its static excess capacity.

11. All this assumes that a short-period output in excess of the critical output can profitably be chosen. A Perfect Sticker will (as we have seen) always choose such an output; but if the short-period costs of producing an output as large as this are high, it is hardly conceivable that the firm will behave as a Perfect Sticker. If such an entrepreneur pays any attention at all to short-period costs, he will always be looking over his shoulder to see if this *strong* policy is worth while. Now it is most unlikely that it will be worth while unless it can be maintained and carried through; thus it will not be worth while unless the long-period output which is being planned for is itself greater than the critical output. In the case of a firm which is operating under strongly diminishing costs, this latter condition is perhaps not unlikely to be satisfied—so that the balance is likely to swing in favour of a *strong* policy. But when costs are falling less steeply, or when the firm has more of the Snatcher mentality, the balance may swing the other way.

But whichever way the balance swings, the general results of our study of oligopolistic expectations go far to confirm those which we got in the polypolistic case. Such firms as are Stickers by disposition and such as have strongly diminishing costs are likely to have a bias in favour of a *strong* policy of excluding competitors; they will accordingly act as if their expectations were no more than mildly polypolistic. Firms whose costs are falling less steeply will be unable to afford so strong a policy; if they are Snatchers by disposition, they will not care to attempt it, even if they could afford it They will thus act as if their expectations were strongly polypolistic. We are thus brought to a classification of firms which appears very similar to that which we reached in the polypolistic case. For firms which belong to the second of these types will behave in a manner which is broadly similar to that which we associated with the marginal firms of our former model; while the first type has the same sort of correspondence with the steady intramarginal firms.

12. I have so far maintained the assumption that the general state of demand for the products of the industry remains stationary, and is expected to remain stationary, throughout the whole story; the various states of expectation which we have been considering differ solely with respect to the expected intensity of competition. There would, I think, be no difficulty in superimposing upon this analysis the consequences of alternative hypotheses about the growth, and expectations of the growth, of this general demand. But most of what has to be said under that head is no way specially concerned with the structure of the market under consideration; it is not at all specially concerned with *imperfect* competition.

If entrepreneurs expect a future expansion of demand, they will install larger plants than they need in the short run; and during the short run these plants will not be used to capacity. In a complete statement of the theory here adumbrated places would have to be found for a number of elementary points of this sort. But I do not think that I need trouble to lay them out here at the tail of this article.

There is, however, one matter which is excluded by the 'static' assumptions which we have maintained, and which does unquestionably belong to the problems we have been discussing. We have been obliged, by our assumption of a stationary state of demand for the products of the industry, to imply that the demand which the firm can expect in the open period is never greater than the demand which it can expect in the closed period—that the one differs from the other solely because of the (possible) arrival of competitors in the latter case. This is evidently unrealistic.[1] Now this point is not fully met by the admission that the theory needs to be extended to meet the case of a general growth in demand. For there remains the possibility (which links on to our discussion of oligopolistic expectations) that the firm itself may not merely carry a *strong* policy to the point of maintaining its close-period market in the open period, but may endeavour to build up its goodwill (by moderation in price-fixing, and also by advertisement and other selling expenditure) to the extent of expanding its market in the further future beyond what it can attain in the near future. Evidently this may happen without an expansion in the *volume* or *scale* of demand for the products of the industry; but it does imply that the firm can influence the *character* of the demand with which it is faced. If such influence is possible (goes on being possible), and if costs are falling in such a way as to make the strong policy possible also, then the firm which is most successful in pursuing the strong policy must come to dominate the industry, so long as the volume of demand for the product as a whole is not increasing. This, however, is in its turn a trite conclusion; I mention it merely in order to show how it links up with the general discussion of this paper.

What I have tried to discuss in this paper is the old story of Imperfect Competition with Free Entry; not Monopoly, nor the Tendency to Monopoly. Some would hold that Imperfect Competition, in this sense, is almost a thing of the past; but even if that were so, it would remain worth study, because it is more nearly an effective alternative to existing forms of industrial organization than the Perfect Competition which preceded it as a theoretical model. The view of it which I get is related to that put forward (in 1953) by Mr. Harrod (and of course owes its origin to Mr. Harrod's work); but it is not identically the same. I agree with Mr.

[1] As Mr. Harrod rightly insisted when he saw a first draft of this paper.

Harrod that there has been too much fuss about static excess capacity; while dynamic excess capacity (very real trouble though it often is) has nothing special to do with imperfect competition. Yet, having said that, I must add that more seems to me to be left of the classical doctrine about the 'wastes of competition' than he seems to have come to allow. Where I still find such wastes are in the disincentive to large-scale investment in an unassured market, and in the incentive to small-scale 'snatching'—not in anything which could be detectable in the regular organization of the ordinary successful, or moderately successful, firm. But these wastes are quite genuine wastes; the fact they have been so generally noticed by practical men does not mean that they are beneath the attention of economic theorists.

Whether these evils are greater than the evils of monopoly, or near-monopoly, is a question on which it is still permissible to have more than one opinion. Personally, I do have more than one opinion. Circumstances alter cases.

ALL SOULS COLLEGE,
OXFORD.

CONVERTIBILITY

By J. R. SARGENT

THERE is no doubt that the best way to neutralize a disturbing idea is to subject it to careful analysis. Convertibility, for good or ill, seems at present to have succumbed to the treatment. The march towards freedom on this front degenerated rapidly into a confused council of war when it was learnt from those familiar with the terrain that the advance could take not one but many possible routes. At least these different approaches have the common characteristic that whichever is followed the freedom with which sterling can be converted into other currencies will be increased; 'an effective convertibility must give some additional rights to someone'.[1] But whether this additional freedom is to apply to residents of the Sterling Area or to non-residents, to individuals or to central banks, to sterling accumulated in the past or accumulating in the future, for purposes of current or of capital transactions; whether greater freedom to exchange sterling is to involve measures increasing the difficulty of obtaining it; and whether the rate at which it can be converted is to be fixed or variable —the very asking of these questions[2] reveals that a large number of brands of convertibility are logically possible. For the sake of argument, however, some can be shrugged off and at least one can be laughed off. I shall assume that there is no question of a free rate of exchange nor of any relaxation of control on the availability of sterling for capital transactions. The latter has the sanction of the International Monetary Fund. The former, on the other hand, offends against its basic philosophy and would strike it a lethal blow. It is inconceivable that the United Kingdom would be willing to do this.[3] Moreover, it seems possible that the equilibrium of the exchange market is an unstable one such that fluctuations in the rate of exchange will exaggerate rather than rectify themselves. This is particularly likely in the case of sterling and would make a free rate extremely hazardous.[4] I shall also assume that, according to the precedent of E.P.U., there will be no convertibility without some special arrangement concerning sterling balances outstanding. When E.P.U. was being negotiated the British insisted that sterling accumulated by members at the date of its coming into operation should not involve the United Kingdom

[1] A. C. L. Day, 'Convertibility and E.P.U.', *Oxford University Institute of Statistics Bulletin*, May 1953.

[2] See Professor J. E. Meade, *Three Banks Review*, Sept. 1953.

[3] There must have been a funereal atmosphere, redolent of a congress of undertakers, at the 1953 Conference of the I.M.F. if the *Banker* (Oct. 1953) correctly reports 'a strong and growing conviction' that the sterling rate must eventually be set free. The conviction is shared by Professor Meade.

[4] Cf. G. Stuvel, *The Exchange Stability Problem.*

in loss of gold. There is no reason to suppose that a different attitude would be adopted now. Consequently, if there are dangers from convertibility, it will be those arising from loosening the reins on sterling accumulating after the date of its introduction that will call for attention.

Thus far we have reduced the possibilities to this: that there will be greater freedom to convert for purposes of current transactions sterling which accumulates after a certain date. Whether this greater freedom will be extended both to residents and to non-residents of the United Kingdom is bound up with the question whether it will be accompanied by changes in commercial policy. If, for example, there are no such changes and Quantitative Restrictions are to remain of predominant importance in guarding the balance of payments, then it is hardly sensible to relax Exchange Control on the current transactions of residents. For the payments which residents want to make abroad are regulated partly by Quantitative Restrictions and partly, in the case of invisible transactions for example, by Exchange Control. To relax the latter while maintaining the former would amount, among other things, to discriminating in favour of invisible payments. There is no obvious economic sense in this.[1] Greater convertibility for residents is therefore likely only in the context of substantial changes in commercial policy. The complementary functions of Exchange Control and of Quantitative Restrictions as regards their effect on residents mean that they must fade or die together. It might be intended in future to use a tariff rather than methods of direct control to keep payments abroad in check; and this would imply the disappearance both of Exchange Control on current transactions and of Quantitative Restrictions, both for residents and for non-residents. Without such changes non-residents are likely to be the only beneficiaries of any increase in the convertibility of sterling.

A case could undoubtedly be made for the introduction of a tariff high enough to enable Quantitative Restrictions to be dispensed with and Exchange Control on current transactions to cease for residents and non-residents alike. Especially cogent is the argument that a tariff, provided it is applied at a uniform *ad valorem* rate over all imports, may be more efficient than administrative controls in selecting those imports for which home-produced substitutes can be found at least additional cost. It may be fanciful to interpret the recent British application to GATT, which sought permission to raise certain tariffs without breaking the rule of duty-free entry for Empire products, as a *ballon d'essai* in this direction.

[1] There would be even less in a policy which relaxed Exchange Control for residents while stiffening Quantitative Restrictions. Or would this perhaps be an infinitely subtle and GATT-proof means of increasing discrimination against the dollar, it being thought that payments controlled by Quantitative Restrictions tended, pound for pound, to have a larger dollar content than those regulated by Exchange Control?

Certainly, however, the reception of the British proposal and the procedural shackles hung upon its translation into practice clearly show that it would be difficult to expand it into a general policy. Moreover, any tariff which replaced Quantitative Restrictions and Exchange Control would have to discriminate against the dollar; to discriminate by tariff is less acceptable to GATT than to discriminate by Quantitative Restrictions. The latter is allowed temporarily for the purpose of rectifying a disequilibrium in the balance of payments, the former not at all.[1] Consequently if Britain were to abandon Quantitative Restrictions in favour of a higher tariff she would lose some of her present power to discriminate against the dollar. It therefore seems likely that substantial changes in commercial policy such as might release residents of the United Kingdom from direct controls, even if they are not a mirage, lie a long way off. If anything is to be done in the near future it can only be the grant of greater convertibility to non-residents.[2]

There are some non-residents whose rights of convertibility cannot be extended because they are complete already. This is true of residents of the Dollar Area. It also applies to overseas members of the Sterling Area with the exception of South Africa. South Africa provides her own hard currency requirements, and if she accumulates sterling can convert it into gold only in the negative sense that she may thereby be spared from purchasing sterling by sales of gold. All other members of the Sterling Area have free access to the Dollar Pool. The readiness with which they resort to it to have their sterling converted into dollars is a matter for themselves alone. In fact it is checked by the various controls operated by the members themselves; but there is nothing that the United Kingdom can do to give them additional rights to draw on it.[3] It must therefore be countries outside the Dollar and Sterling Areas that will be the beneficiaries of any increased convertibility of sterling.

[1] Except in the case of existing margins of preference. In the British case these discriminate against the dollar only in the general sense that they favour the Empire, and certainly not to the same extent as Quantitative Restrictions.

[2] As argued above, this would be sterling arising out of direct current transactions after a given date. References to 'convertibility' in what follows are to be interpreted in this sense.

[3] This argument, while formally correct, may be misleading. The convertibility enjoyed by the overseas members of the Sterling Area would be more effective if they were able to relax the intensity with which they discriminate against the dollar. For the present degree of discrimination the United Kingdom is directly responsible as far as the colonies are concerned and indirectly as regards the independent members of the Sterling Area. If the United Kingdom is in a mood to toy with convertibility she might do well to give priority to relaxing the discrimination which she enforces or fosters in the overseas Sterling Area. This would not affect the latter countries' rights of convertibility as such, but it would give more substance to the shadow. Without some knowledge of the extent to which overseas members of the Sterling Area do, in fact, discriminate—an unexplored territory—it is difficult to guess the size of the results which would follow from their ceasing, or partly ceasing, to do so. This article, therefore, evades this issue and interprets convertibility in its strict and formal sense.

These countries already possess considerable facilities for the transfer and conversion of sterling for direct current transactions.[1] The most obvious objective of any 'move towards convertibility' would be to widen these existing rights to their limit; namely to allow all non-residents to convert sterling into dollars for direct current transactions. The effects of such a course have been analysed elsewhere;[2] here it will be enough to recapitulate those which are most patent. First there would be a danger that countries whose sterling had newly become convertible would attempt by discriminating against it to accumulate balances for conversion into dollars. There would seem to be little point in doing this unless it were feared that convertibility could not last long. But to dispel such fears it would need to be launched upon a substantial cushion of gold and dollars capable of absorbing large speculative movements of this kind. To provide such a stabilization fund the present reserves of the Sterling Area do not seem to be adequate; and offers from elsewhere have hardly been embarrassingly abundant.[3] A second possibility is that, for sound economic reasons, countries hitherto forced by the inconvertibility of sterling to discriminate in its favour will shift to the Dollar Area for supplies, converting the sterling saved into dollars for the purpose. This would compel the United Kingdom to increase its own discriminatory restrictions on dollar expenditure. In this case convertibility would worsen the terms of trade of the United Kingdom and improve those of its competitors. This tendency to shift sources of supply would not be confined to countries outside E.P.U. The surplus which an E.P.U. member can at present accumulate in the Union by reducing its imports from the Sterling Area is never entirely convertible into gold. The practice of the Union has been never to allow creditors to receive 100 per cent. gold in settlement of their monthly surpluses even when they have passed the limit of their quotas. Consequently an E.P.U. member deciding to switch its source of supply away from the Union by £10 million will only receive gold to the extent of something less than the full £10 million of its resulting surplus in the Union. Depending on the relative prices of dollar substitutes, the gold received may or may not be enough to provide an equivalent volume of supplies from dollar sources. If, however, sterling were convertible, an E.P.U. member could

[1] 'Direct current transactions' are defined as follows: 'payments . . . which are made by a principal resident in the country from which payment is made and which relate solely to goods (other than gold bullion, gold coin, or gold either in semi-manufactured or fully manufactured form) imported into and for use or consumption in that country and originating in the country to which payment is made, or to services rendered to residents of the former country by residents of the latter country'. The word 'direct' is intended to exclude 'commodity shunting' operations; for example, the use of sterling by a Dutch merchant to buy Australian wool for sale to an American importer.

[2] See, for instance, the articles by A. C. L. Day and Professor Meade referred to in footnotes to page 55.

[3] See the *Banker*, Oct. 1953.

reduce its imports from the United Kingdom by £10 million[1] and, instead of throwing its resulting surplus into the clearing, could convert the whole into dollars. In this case it could shift its source of supply to the Dollar Area with no net expenditure of dollars, whereas previously some extra dollar outlay might be required, depending on the member's cumulative position and the dollar cost of replacing the original supplies.

Even if there were no shifts in sources of supply and no deterioration in the trade balance of the United Kingdom and the rest of the Sterling Area as a result of convertibility, there would be dangers arising from the low level of their gold and dollar reserves. Although on an average the supply of sterling might be kept equal to the demand, small temporary increases in the Sterling Area's overseas expenditure would cause extravagant drafts on the reserves. If imports rose or exports fell, and sterling accumulated in the outside world which the latter exercised its right to convert into dollars, the Sterling Area would have to resort to import restrictions more quickly and more drastically than would be seen to be justified in the long run. For when reserves are short they have to be defended even against passing attacks which will come to nothing in the long run; and in their defence quantitative and other restrictions will be imposed, and will create disturbances in the normal patterns of trade, which will later be seen to have been unnecessary if only reserves had been large enough to absorb the temporary fluctuation. Thus convertibility on the basis of slender reserves may involve a loss of economic efficiency, by diverting trade from its normal channels, fully as much as inconvertibility.

Even if the desire of non-residents to shift towards dollar sources of supply is weak, and even if the aggregate supply of sterling in the outside world does not exceed the demand either in the long run or temporarily, there may still be sudden pressure on the gold reserves if there are shifts in the pattern of trade balances. If country *A* imports less from the United Kingdom and accumulates a sterling balance, while country *B* imports more and is short of an equivalent amount, there is a risk that *A*'s balance will be converted into dollars while *B*'s deficit will be met by restrictions on sterling imports. Although the United Kingdom began with neither a surplus nor a deficit, a mere shift in the pattern of trade balances will have saddled it with a deficit paid for entirely in gold. This might be avoided in two ways. First, it might so happen that *A* had concurrently developed a deficit with *B* which it had to use its surplus sterling to discharge. Secondly, the danger outlined would not exist if there were general

[1] In so far as it was able to make use of the 'sterling option', i.e. the provision that surpluses may be held in sterling (or any other E.P.U. currency) instead of E.P.U. units of account, it would not matter whether imports were reduced from the U.K. in particular or from the Union as a whole. If the latter, the resulting surplus could be held in sterling and this converted into dollars.

convertibility. In this case the United Kingdom's loss of gold to *A* would be recouped from *B*. In the absence of this there is the general danger that sterling surpluses will be converted while sterling deficits will be met by other means.

The first of these effects—that convertibility would make possible desired shifts to dollar sources of supply—while endangering the balance of payments and the competitive position of the United Kingdom, would be justified in the eyes of the 'convertibility-as-a-spur' school of thought. This school is now presided over by the International Monetary Fund, which declares in its latest Annual Report:

> It is now more widely realized that the elimination of the dollar shortage requires adjustments in the pattern of international trade and payments to which an effective movement in the direction of convertibility might itself make a significant contribution. . . . Steps towards convertibility would support efforts directed at overcoming the shortage of dollars, both by giving countries stronger inducements to undertake and maintain monetary fiscal and budgetary policies that would prevent over-expansion of domestic demand and by removing some of the rigidities in the international payments system that impede the re-establishment of a satisfactory pattern of world trade.

The Fund, however, goes on to recognize that the benefits of convertibility would be qualified if it involved contraction in other directions. This is precisely what the second and third of the effects analysed above would do. After the United Kingdom had been forced into dollar equilibrium by the necessity of meeting demands for convertibility from those shifting to dollar sources of supply, there would still remain the danger that temporary fluctuations in the supply of sterling, which might themselves be small in relation to the total supply, would be large in their effect, via convertibility, on the gold reserves and therefore necessitate restrictions out of proportion to their real significance. Furthermore, such restrictions might also be called for when there was no excess supply of sterling (even of a temporary nature), but simply as a result of a shift in the pattern of trade balances.

Full convertibility, therefore, would bristle with difficulties and dangers for the Sterling Area. The movement towards it, conceived at the Commonwealth Conference of January 1953, has now ground to a halt less from a lack of resolution on the part of the United Kingdom authorities to face them than from anxiety in the rest of the world concerning what would happen when they did. It has been the general fear, especially within E.P.U. where the memory of the United Kingdom's retreat from liberalization at the end of 1951 is strong, that, in much the same way as the gold pound of 1925, so the convertible pound of 1953 would be made a scarce one. Hence the cool reception of the project and its tendency recently to wilt in the arms of its sponsors. But if full convertibility is no longer a live issue, it might still be profitable to travel some way along the

road towards it. It might be possible to increase the transferability of non-resident sterling as far as possible without granting convertibility into dollars; or in other words to make sterling convertible for non-residents into the currencies of all countries except those of the Dollar Area. At present these countries are divided into two main classes according to the degree of convertibility which they enjoy; namely, those who are members of the Transferable Accounts Area, and those who are designated Bilateral Accounts.[1] Members of both these areas can freely transfer sterling which they may accumulate for purposes of direct current transactions to any member of the Sterling Area; but members of the Transferable Accounts Area are allowed to transfer it to each other. Countries of the Bilateral Group do not share this latter privilege; and a resident of one of these, if he does not wish to spend his sterling in the Sterling Area, can only use it to pay a fellow countryman. This classification, however, is cut across by the existence of E.P.U. There are E.P.U. members to be found in both groups. Within E.P.U. sterling is freely transferable; therefore there are some members of the Bilateral Group able to transfer sterling both among themselves and outside the group to other members of E.P.U., and some members of the Transferable Accounts Group for whom the transferability of sterling is not limited to making payments to other members of the group, but is extended for making payments outside it to other members of E.P.U. The freedom with which countries outside the Sterling and Dollar Areas can transfer sterling among themselves is summed up synoptically in Table I. The rigid outlines of this picture are softened in practice by the existence of 'administrative transferability'. Transfers not freely allowed —for example, between two Bilateral countries or between a Bilateral and a Transferable—may be made if the Bank of England consents, and have in fact in recent years been almost as important as those effected automatically.[2] In the case of Japan, a Bilateral country, the United Kingdom specifically committed itself, during the currency of the 1951 agreement,[3] to allow any transfer of sterling by Japan to any member of the Transferable Accounts Area for purposes of direct current transactions.[4] In the

[1] There is a small residual group of countries with whom it is unimportant or impossible to conclude definitive payments arrangements. For the classification of countries into those with different types of account, as at 31 Dec. 1952, see Fourth Annual Report on Exchange Restrictions, published by the International Monetary Fund. To the date of writing (31 Dec. 1953), there have been no changes in this classification.

[2] Report of the Bank of England for year ended 28 Feb. 1953 (Cmd. 8911).

[3] It expired on 31 Dec. 1953. Negotiations were in progress at the date of writing (31 Dec. 1953) for a new agreement, but the main point at issue was not Japan's freedom to transfer sterling, but her extreme shortage of it.

[4] This seems to be unique among the United Kingdom's agreements with Bilateral countries (see Exchange of Notes subsequent to Payments Agreement of 31 Aug. 1951; Cmd. 8602). It is evidently not intended, however, to countenance an excessively multilateral use of sterling, since the agreement also provides that the two countries will take 'all reasonable means' to prevent a chronic imbalance of payments in either direction.

case of Iran a somewhat similar pledge has been given. In September 1951 Iran was removed from the Transferable Accounts Area to the Bilateral Area; but it was stated by the Treasury that the additional power thus obtained over Iranian sterling would normally be exercised in such a way as to allow all transactions except conversion into dollars. The United Kingdom has also committed itself to receiving into the Transferable Accounts Area any E.P.U. member not already included in it. Since E.P.U. began, however, only Greece, Austria, Denmark, and Western Germany have made use of the offer. Those E.P.U. countries still in the Bilateral Group are there of their own choice.

TABLE I

Transferability of sterling for direct current transactions among non-sterling, non-dollar countries

From / *To*	(1) *Members of Transferable Accounts Area* AND E.P.U.	(2) *Members of Transferable Accounts Area but* NOT E.P.U.	(3) *Members of Bilateral Group* AND E.P.U.	(4) *Members of Bilateral Group but* NOT E.P.U.
Members of Transferable Accounts Area AND E.P.U.	Yes	Yes	Yes	No
Members of Transferable Accounts Area but NOT E.P.U. . .	Yes	Yes	No	No
Members of Bilateral Group AND E.P.U. . .	Yes	No	Yes	No
Members of Bilateral Group but NOT E.P.U. .	No	No	No	No

Key to membership of each group:

(1) Austria, Denmark, German Federal Republic, Greece, Italian Monetary Area, Netherlands Monetary Area, Norway, Sweden.
(2) Sudan, Chile, Czechoslovakia, Egypt, Ethiopia, Finland, Poland, Spanish Monetary Area, Thailand, U.S.S.R.
(3) Belgian Monetary Area, French Franc Area, Portuguese Monetary Area, Switzerland, Turkey.
(4) Argentina, Brazil, Bulgaria, China, Hungary, Iran, Israel, Japan, Lebanon, Paraguay, Peru, Rumania, Syria, Tangier, Uruguay, Vatican City, Yugoslavia.

Hitherto we have outlined the existing rights of countries outside the Sterling and Dollar Areas to transfer sterling among themselves. These rights are not exhaustive of all possibilities, as can be seen from the number of 'No' entries in Table I. Without, therefore, going all the way to full convertibility, it would be possible to turn the 'No' entries of Table I into

'Yes'; to complete the freedom of countries outside the Dollar and Sterling Areas to transfer sterling in direct current transactions to each other. The effect of this would be to eliminate a number of the discriminatory features of the present system which seem to have little economic justification, while retaining the power of Exchange Control to enforce discrimination against the dollar. At present, so far as they may accumulate sterling and may wish to spend it, E.P.U. countries in the Transferable Group are forced to discriminate against the dollar and against the Bilateral countries outside E.P.U., and in favour of sterling, of E.P.U. currencies, and of other members of the Transferable Accounts Area. Countries not members of E.P.U. but in the Transferable Accounts Area must discriminate against all in the Bilateral Group and the dollar, and in favour of sterling and of other members of the Transferable Accounts Area. Countries in E.P.U. but in the Bilateral Group must discriminate in favour of sterling and of other E.P.U. currencies, and against all others, while countries in the Bilateral Group but not members of E.P.U. must discriminate in favour of sterling and against all other currencies. These curious patterns of discrimination may be relaxed by administrative transferability; and the extent to which they divert trade away from those channels which are most economic may not be of great importance compared with the effect of trading policies conducted by the countries themselves. But if sterling were to be made freely transferable by non-residents for direct current transactions anywhere in the world except the Dollar Area, at least some transactions would become possible which are now thwarted and in consequence involve the finding of uneconomic substitutes. This would mean bringing all countries outside the Sterling and Dollar Areas into an enlarged Transferable Accounts Area. We must see what dangers it would involve.

These dangers arise from the fact that some countries outside the Dollar and Sterling Areas have rights of convertibility into gold or dollars. If sterling accumulates in E.P.U. the extent to which it will require to be settled in gold will depend upon the cumulative accounting position of the United Kingdom, and can reach 100 per cent. Outside E.P.U., however, the only country with convertibility rights at present is Argentina.[1] This arrangement dates from the agreement of April 1951,[2] but was put in its present form by that of December 1952.[3] The latter provides that if Argentine Sterling Balances rise above £20 million, the balance shall be available, after repaying certain advances from the United Kingdom

[1] The special privileges of Iran were withdrawn in September 1951. Egypt, under the Sterling Balances Agreement of July 1951 (Cmd. 8337), was sold £14 million in dollars and given the right to pay sterling for imports of petrol up to £11 million annually in each of the years 1951–60. But no permanent arrangement for the conversion of Egyptian sterling into dollars was arrived at.

[2] Cmd. 8231.

[3] Cmd. 8744.

which may be outstanding, for transfer to American accounts or conversion into dollars. These dollars must be sold back, however, if the sterling balances fall below £20 million. This means that some time after the beginning of the agreement Argentina may have become a net gainer of dollars from the United Kingdom, but the United Kingdom will never have become a net gainer of dollars from Argentina. The members of E.P.U. and Argentina are the only countries whose sterling the United Kingdom is publicly committed to allowing to be converted into gold or dollars. There may, however, be others from time to time whose bargaining power waxes owing to an accumulation of sterling and enables them to demand similar treatment. The danger of widening the transferability of sterling would be that it would increase the number of sources from which sterling might flow into balances requiring conversion into gold or dollars.

Let us be more specific. We will consider each line of Table I in turn and assess the effects of crossing out the 'No' and substituting 'Yes' in each column where this can be still done. First, this would involve the danger that countries in the Bilateral Area not in E.P.U. could add to the flood of sterling into E.P.U. Of the countries in this group, however, the most important trading nations are Argentina, Brazil, and Japan. Japan, although not a member of the Transferable Accounts Area, has already been assured of this facility.[1] Brazil is at present a heavy sterling debtor. Under the settlement of 2 October 1953 she is to pay off £6 million each year, with interest of 3½ per cent. and additional capital payments according to a sliding-scale percentage of her sterling receipts. Consequently she is unlikely for some years to find herself with surplus sterling which might make use of greater transferability. In the case of Argentina, the extent to which any transfer of sterling she might make to E.P.U. would affect the United Kingdom's gold reserves would depend upon the level of her sterling balances. If these exceeded £20 million, the United Kingdom would either have to convert the excess into dollars or allow its transfer to E.P.U. and meet the consequences there. Provided that the United Kingdom had not exhausted its E.P.U. quota and was not therefore meeting its deficits entirely in gold, the result of allowing this degree of transferability to Argentina would be a net saving of dollars. If, on the other hand, Argentine sterling balances were below £20 million, the effect of the transfer might be to increase the amount of dollars that the United Kingdom could claim back from Argentina; and again, the dollars lost to E.P.U. could only be as great as those regained from Argentina if the United Kingdom had exhausted its E.P.U. quota. But under the agreement the United Kingdom can only claim back dollars previously sold; and if there were

[1] See above, p. 61.

few of these, the transfer of Argentine sterling to E.P.U. would be more likely to involve a loss of gold. All in all, however, as regards the three major trading nations of this group, the risks of allowing them to transfer sterling to E.P.U. do not seem great.

Consider now the second line of Table I. If the 'No's' here become 'Yes' there would be a larger number of sources from which sterling could trickle into the accounts of members of the Transferable Accounts Area outside E.P.U. Among these are none enjoying rights of convertibility, although Egypt is capable of successful agitation. On the face of it, therefore, greater transferability to these countries from the Bilateral Group would diminish the risk of loss of gold; for it might attract sterling away from Argentina and away from those E.P.U. members who are in the Bilateral Group. As far as the latter are concerned this effect would be small. They have already been offered membership of the Transferable Accounts Area; the fact that they have shown no eagerness to accept indicates that they are not labouring under many unfulfilled ambitions to transfer sterling to outlying parts of the Transferable Accounts Area. Indeed, since they comprise three large creditors (Belgium, Portugal, and Switzerland) and two large debtors (France and Turkey), in E.P.U., they all have strong reason to throw their sterling into the clearing and gain or save an almost equivalent amount of gold.

Thirdly, line 3. Here we might have sterling entering E.P.U., not only from the sources discussed in considering the effects of liberating line 1, but also from members of the Transferable Accounts Area outside E.P.U. The former did not seem likely to be dangerous; the latter, however, may. For none of the countries of the Transferable Accounts Area who do not take part in E.P.U. require gold settlement. Hence, if greater freedom allows them to transfer sterling to E.P.U. where some gold settlement is inevitable, loss of gold is bound to result from such transfers as are made. The only exception will be if Egypt, for example, having accumulated sterling some part of which she is pressing to have converted into gold, is content to be permitted to transfer it to Belgium; and if the United Kingdom's cumulative position in E.P.U. involves only a small gold settlement in respect of the extra sterling to be cleared.

Finally, line 4. What if there is greater freedom for sterling payments to those members of the Bilateral Group who are not members of E.P.U.? The risk here is primarily that of sterling accumulating in Argentine accounts. On the other hand is the possibility that greater freedom in this direction may induce some members of E.P.U. to employ their sterling in, for example, Japan or Brazil where no gold settlement is required. This possibility, as we have seen, is unlikely to excite those E.P.U. members who have been offered but have not accepted Transferable Account

facilities. Probably it will not excite the others either; for it seems unlikely that transfers of sterling have ever been denied administratively, even if they have not been permissible automatically, in cases where they would have relieved the loss of gold to E.P.U.

This is no doubt of general application. If non-residents have wanted to transfer sterling from an account which would otherwise have required gold settlement to one which would not, we can assume that they have been permitted to do so by administrative action wherever they have not enjoyed the automatic right. Consequently we cannot hope that greater transferability for non-resident sterling will enable any to escape the fate of being turned into gold. On the other hand, it will increase the ease with which it can trickle into E.P.U., into Argentine account, and into other accounts where the accumulation is such as to put pressure on the authorities to allow its conversion.

Against this risk to the gold reserves of the Sterling Area must be set the advantages to the rest of the world (except the Dollar Area, which enjoys them already) of the removal of barriers to the expenditure of sterling which may have forced them towards costly substitutes for the imports they would have liked. The United Kingdom's case for refusing to remove these barriers would be weak indeed if it were merely trying to protect itself from the consequences of profligacy. But this is not the purpose of the barriers. Even if the supply of sterling in the world outside the Sterling Area is, taking one year with another, kept equal to the demand, it is bound to be subject to fluctuations. If suddenly a temporary excess of sterling appears and it is easy for it to find its way into convertible accounts, the risk of disturbance to the normal course of trade is correspondingly great, although before long the excess may vanish. A shift in the pattern of trade, without the appearance of any excess supply of sterling, might have the same disturbing effects. If country *A* develops a sterling surplus and wishes to transfer it to E.P.U. while *B* runs into a deficit and finances it from sterling balances acquired in a previous period, there is a threat to the Sterling Area gold reserves, and this will be the greater the larger number of countries able to transfer their surpluses of sterling to E.P.U. Except in the event that gold reserves can be increased, greater transferability for non-resident sterling may substitute trade which is freer but more subject to disturbance for trade which is more stable although in some cases diverted into curious directions.

There is, however, another danger which would arise if the present Transferable Accounts Area were widened to include all countries outside the Sterling and Dollar Areas. This is the danger that transfers would not be held down to those in respect of direct current transactions. Transferable Accounts are not held by individuals but by banks; and payments of

sterling from an individual resident of one Transferable Account country to an individual resident of another are made, not directly, but via the Transferable Accounts of banks, or, in many cases, of the central banks of the territories concerned. The purpose of this is to ensure that some watch can be kept on sterling transfers in order to see that they are not made to finance capital transactions or to facilitate current transactions merely of an entrepôt nature. Most of the United Kingdom's monetary arrangements run to a standard form; an agreement with Ruritania would, for example, provide:

(*a*) that sterling *at the disposal of residents of Ruritania* shall be freely available for payments to the Sterling Area;

(*b*) that sterling *under the control of the Bank of Ruritania* shall be available for payments in respect of direct current transactions to such countries outside Ruritania and the Sterling Area as may be agreed between the United Kingdom and Ruritania.

The latter clause is common to agreements both with countries having Transferable Accounts and with those that have not. But it suggests that the admission of a country to the Transferable Accounts Area depends upon its willingness to take control of its residents' sterling and see that its use is confined to direct current transactions. When new Transferable Accounts are set up, the Bank of England abandons the supervision of the payments of sterling which they may be used to make, and implicitly commits it to the foreign bank in whose name the Transferable Account facilities have been opened. The degree to which the latter can be relied upon to restrict unauthorized capital transactions, or current transactions of an indirect nature, no doubt varies from country to country. But if a country is regarded as unreliable in this respect, the chances are that it is now outside the Transferable Accounts Area. Consequently to widen the membership of the area would be to widen the risk of sterling accumulating in convertible accounts, since its opportunity for reaching them would not be so easily confined to direct current transactions.

From this analysis it appears that what stands in the way of wider convertibility of sterling in the non-dollar world is the existing degree of its convertibility into dollars. In the strategy of the dollar crisis much might be gained if sterling were made freely available for direct current transactions in any country outside the Dollar Area. For this would help to enforce discrimination against the dollar without implying discrimination in favour of any other currency. Thus the real costs of discrimination would be reduced. For each non-dollar country, to the extent that it might come into possession of sterling, would be able to use it[1] for imports from the

[1] Except in so far as the country might itself prevent it.

cheapest, or in other ways most desirable, source of supply, whereas at present this may be checked by the incomplete transferability of sterling within the non-dollar world. If, however, this transferability is to be made complete, the United Kingdom has to face the risk of widening the roads which lead to convertible accounts. This risk could be accepted readily if there were some modification in the degree of convertibility into dollars required by those in possession of convertible accounts; if, for example, the scale regulating the settlement of surpluses and deficits in E.P.U. were revised in such a way as to reduce the proportion payable and receivable in gold. The Managing Board of E.P.U. was instructed at the O.E.E.C. meeting in Paris at the beginning of November to report, before 31 March 1954, 'on the methods which might be adopted for an orderly transition . . . from the present form of E.P.U. to a wider system of trade payments and credits'.[1] One limb of such a wider system might be the extension of the rights of non-residents (including those outside E.P.U.), with regard to sterling they may accumulate, to the point where it is freely transferable anywhere in the world except the Dollar Area. Whether the members of E.P.U. would be sufficiently attracted by this possibility to agree to lowering the percentage of gold in the settlement of monthly surpluses and deficits is uncertain.[2] Some, indeed, have been pressing for the reverse; and some have not felt attracted by the United Kingdom's offer of membership of the Transferable Accounts Area, with the wider powers that it offers for the disposal of their sterling. Consequently the bait of free transferability of sterling within the non-dollar world might not be strong enough to elicit the *quid pro quo* which would reduce its risks. In this case the sweeping aside of the largely pointless patterns of discrimination illustrated in Table I might be postponed. It should remain, however, an objective considerably more attainable than full convertibility and more certainly beneficial to the world at large in its effects.

[1] See *The Economist*, 7 Nov. 1953.

[2] It would also be necessary to press for modification of the agreement with Argentina in the same sense.

SOME ECONOMIC ASPECTS OF COMPULSORY TRADE UNIONISM

By V. L. ALLEN

In this paper I am concerned with the motives that impel management in industry to accept compulsory trade unionism, with the forms it takes, and with the extent of its practice in British and American industry. Fundamental political issues are involved, so far as the individual worker is concerned, wherever compulsory trade unionism exists, but for the purpose of this paper they can be ignored. My material has been drawn from mixed sources: from trade-union constitutions, collective agreements, trade-union reports; from men whose daily task it is to make trade-union membership compulsory or to solve the problems to which it gives rise; and from my own observations, covering a number of years, as a lay member of a craft union which has tried but failed to use compulsion. I am indebted to those representatives of both sides of industry who have provided me with information.

The efficiency of the firm is influenced, amongst other things, by the quality of its labour force. In the choice of workers consideration normally has to be given to age, training and experience, and standard of health of the workers; in the promotion and dismissal of workers, efficiency, calculated to include work speed, adaptability, quality and knowledge of operations, is normally the decisive criterion. In the past it has been a managerial prerogative to choose, promote, and dismiss workers, and this was largely accepted by trade unions. During recent years, on the grounds that these are not simply economic decisions affecting the firm but welfare decisions also, trade unions have challenged this managerial prerogative and have achieved some degree of success in negotiating working rule agreements to establish norms and procedures for such actions. In some cases management has relinquished the right to choose workers according to its requirements and instead has agreed that the criterion for selection should be membership of a trade union, thus making membership of a trade union compulsory within particular factories or industries.

Types of Agreement

No compulsion can be exercised by trade unions over workers unless employers agree to collaborate. Though this collaboration may be tacit, and is in fact so throughout the major part of British industry where compulsion is present, it is more effective when written into an agreement. Such agreements take two forms, 'union shop' agreements and 'closed shop' agreements.

A 'union shop' agreement normally requires a worker to join a union after he is hired but before a specified time-limit; usually between 25 and 30 days from the time he starts work. Once the worker joins the union he is compelled to remain in it for the duration of his employment at that establishment. There are variations of this agreement. For example, the Scottish Union of Bakers, Biscuit and Bakery Workers rules that 'all workers engaged by an employer shall require to be members of the Union, or to apply for membership of the Union within 30 days of entering such employment, provided that it shall be open to the employer or the worker to seek permission of the National Joint Reference Committee . . . for the exemption of such worker from the terms of this rule, stating the grounds on which exemption is sought'.[1] The workers, under this form of agreement, are given time to consider the rationality of joining a trade union; to assess, that is, how far it is a sound benefit society, whether its conditions will increase or decrease liberty, and to what extent the conditions of employment will improve or deteriorate as a result of trade-union regulations.[2] But compulsion over the worker is exercised if he is refused employment at the end of the time-limit because he will not join the union. The worker's attitude, when faced with this dilemma, will undoubtedly depend on the availability of alternative means of employment. This does not simply mean his chances of getting another job of any kind: many factors have to be taken into consideration, such as the location of the job, relations with management, remuneration, work-group solidarity, and the physical and mental requirements of the job. A skilled worker with apprenticeship training would find it more difficult to obtain comparable employment than an unskilled or semi-skilled worker.

An employer who has bound himself with a 'union shop' agreement would not be able to keep a worker he had found to be efficient, beyond the end of the time-limit, unless that worker agreed to join a trade union. Furthermore, if an efficient worker ceased to be a member of a trade union during the course of his employment then the employer would be obliged to dismiss him. The employer would still be permitted to choose his own labour in the initial stages, but his subsequent control over it would be reduced.

The 'closed shop' agreement is a much more stringent form of regulation of both employers and workers. When it covers skilled craftsmen the trade-union card becomes the key to employment, for workers must already be members of a trade union before they can be considered by an employer. Provided this condition is satisfied the employer retains the

[1] Rule 21; established by the National Conciliation Agreement between the Union and Employers' Organizations on 21 Nov. 1944.

[2] *Industrial Democracy*, by S. and B. Webb, p. xxx.

right to choose his own labour. In certain trades not recognized as being skilled, membership of a trade union is a condition of employment. The employer can choose his labour, as in the case of 'union shop' agreements, but workers must agree to be members of appropriate trade unions and to obey their regulations immediately they commence work. The extreme case is where an employer uses a trade union as the sole agency for placing workers in his employment. This is the practice of the Federation of London Wholesale Newspaper Distributors, which employs 2,600 workers on distribution and obtains all of its labour through the National Union of Printing, Bookbinding and Paper Workers.[1]

'Closed shop' agreements may or may not stipulate membership of one particular union which claims exclusive bargaining rights.

Policy and Practice

Because of the diverse nature of British trade unions there are bound to be different methods used to achieve an effective enrolment of all workers into trade unions. Some trade unions have written into their constitutions the right of members to refuse to work with non-unionists. The operation of such rules depends, of course, upon the state of the labour market. The Amalgamated Society of Boot and Shoe Makers and Repairers rule that 'When three-fourths of the men working for a shop are Members of this Society, our Members can, with the approval of the Executive Council, withdraw their labour and be supported in the event of a Section's or Investigator's negotiations failing to induce the employer to employ only Society men'.[2] The problem is placed squarely in the hands of the branch by the National Union of Printing, Bookbinding and Paper Workers, which states 'If any person, not being a member, be put to work at the bench . . . the members of the Union shall, on the advice and authority of the Branch, take such steps as shall seem judicious, either for his removal or entry into the Union, and should these measures fail, the Branch shall endeavour to deal directly with the employer.'[3] Only when this step is unsuccessful is the matter to be referred to the Executive Council of the union. Even the Iron and Steel Trades Confederation, noted for its conciliatory approach to labour problems, has consistently opposed the employment of non-unionists.[4] In his presidential address at the Trades Union Annual Congress, 1952, Mr. Deakin stated forcibly: 'We challenge the suggestion that workers should be free to join a union or stay outside, and shall continue to assert with the utmost vigour the right of trade unionists to say whether or not they will work alongside

[1] Cmd. 8607, p. 29.
[2] Rule 28, clause 6. Passed at the General Conference of the Union, 1946.
[3] Rule 14, clause 7 of the General Rules, 1948.
[4] *Men of Steel*, by Sir Arthur Pugh, p. 411.

non-unionists.'[1] A number of unions, all in the printing industry, actually prohibit their members from working with non-unionists.

Regardless of their trade-union rules, many trade unionists are unwilling to work with non-unionists. It has been noted in a study of strikes that strikes over trade-union principles are changing their character and are no longer strikes for the right to organize or for recognition so much as for the right to impose compulsory membership of trade unions on workers.[2] There are many unrecorded 'frontier' battles that are waged over compulsory trade unionism and last for only an hour or so on each occasion. In one large motor-car factory it has been the industrial policy of the shop stewards' committee to organize 'snap' strikes in departments in which non-unionists were employed. The practice was to hold a card check and then inform the management that after one week industrial action would be taken in all the departments in which non-unionists were employed. Trade-union organization in this factory, despite this intensive campaigning, is still not fully comprehensive.

The response of employers to this type of pressure varies considerably. In 1946 the London Passenger Transport Board maintained twelve New Cross tramwaymen with pay for three months without requiring them to work because the rest of the depot refused to work with them.[3] Prior to the First World War industrial unionists directed their propaganda towards the elimination of non-unionists and inspired a number of strikes in the building industry. The London Master Builders retaliated by calling a lock-out and refused to admit any workers to their shops until they had signed a 'document' declaring their willingness to work with non-trade-union members. The lock-out failed because 'even non-unionists, with irrational loyalty, were refusing to sign a document promising not to strike against themselves'.[4] But neither did the trade unions involved obtain their demand.

Many employers are prepared to agree tacitly to the exercise of compulsory membership of trade unions. Some of them prefer to encourage workers to join unions in order to avoid trouble in the workshop if trade-union members there feel deeply about the presence of non-unionists.[5] Others, like the Civil Service, prefer their workers to be members of trade unions because 'the existence of fully representative associations not only promotes good staff relations but is essential to effective negotiations on conditions of service'.[6] Some employers like to feel that the workers'

[1] *T.U.C. Annual Report*, 1952, p. 80.
[2] *Strikes—A Study in Industrial Conflict*, by K. G. J. C. Knowles, p. 237.
[3] *Labour Relations in London Transport*, by H. A. Clegg, p. 129.
[4] *The Builders' History*, by R. W. Postgate, p. 416.
[5] Cf. *Productivity and Trade Unions*, by F. Zweig, p. 16.
[6] *Staff Relations in the Civil Service*, p. 3.

representatives with whom they negotiate can speak with unquestioned authority for workers in the trade or category of industry they represent. However, even employers who are fully sympathetic to trade-union aims, and who are willing to exercise persuasion to the point of coercion, might be unwilling to dismiss an efficient worker who was expelled from his union.

Compulsory Trade Unionism in Different Industries

The attitude of employers to compulsory trade unionism is determined mainly by factors affecting the efficiency of the firm. It is true that in certain local government authorities and the Co-operative Movement, to a large extent, economic efficiency is placed second to political motives. 'Closed shop' conditions in the Co-operative Movement are not the result of collective bargaining but are, in effect, employer-imposed conditions. The procedure is for Co-operative members' meetings or meetings of boards of management to pass resolutions to employ trade-union labour only. The motives for the resolutions appear to be derived from the theoretical identity between the aims of the Co-operative Movement and trade unions. Out of a total of 1,287 Co-operative Societies engaged in retail and wholesale distribution and in production,[1] only 553 societies have passed such a resolution.[2] These societies, however, employ the majority of the workers employed in the movement, and it is probable that other societies employ trade-union labour only without being pressed by resolution to do so.

Compulsory trade unionism may be preferable to work stoppages resulting from friction between trade unionists and non-unionists in industries which have high overhead costs relative to labour costs, and in which the costs of recharging furnaces and resetting and retooling are high. Mainly for this reason workers in the iron and steel industry have been compelled to join trade unions. In all the large continuous-production industries the maintenance of industrial peace is a more important cost factor than the facility to select workers according to their ability and to regulate their engagement with the firm. These industries have a relatively high degree of mechanization; production processes are broken down into small, simple operations that can easily be performed by semi-skilled and unskilled labour; and the working-pace is largely controlled by the speed of the machines. The efficiency of labour under these conditions is relatively unimportant. It is quite possible, then, to find some form of 'union shop' in these industries.

[1] *Ministry of Labour Gazette*, p. 377, Nov. 1952.

[2] A leaflet issued by the Union of Shop, Distributive and Allied Workers entitled *Does your Co-operative Society employ T.U. Labour.*

Even where labour costs form a high proportion of total costs of production the management may gain in the long run from compelling all of its employees to be trade-union members provided that (i) workers are compelled to work to a scheduled pace, (ii) the tasks require men of no more than average ability, and (iii) interrupted production or provision of service is relatively costly. This explains why the London Passenger Transport Board has been willing to operate the 'closed shop' agreement with the Transport and General Workers' Union since 1946. Provided a driver has passed his driving-test and the company has an efficient schedule system, it can be reasonably assumed that compulsory trade unionism would have no adverse effects on the efficiency of the undertaking. Before 1946 the London Passenger Transport Board was covered by Section 6 (1) of the 1927 Trade Disputes and Trade Unions Act which made it unlawful for any local or other public authority to make membership of a trade union a condition of employment.[1]

Port employers are placed in a similar, though more exaggerated, position to that of the London Passenger Transport Board. Dock workers form closely-knit communities with an acute sense of solidarity irrespective of the nature of their grievances. Industrial peace, therefore, always lies uneasily, even when obvious points of friction have been removed. From the employers' point of view it is clearly undesirable to have stoppages of work for two main reasons: (*a*) a stoppage would dislocate the shipping schedules and cause congestion in the ports; (*b*) the port transport industry is a public service of the utmost importance. Since the Shaw Award of 1920 the employers have generally recognized this, though not always sufficiently or successfully. By accepting compulsory trade unionism port employers have removed one obvious point of friction. The employers and dock workers are controlled by the Dock Workers (Regulation of Employment) Order, 1947. The recruitment of workers and their allocation to employers are governed by the provisions of the scheme administered by the National Dock Labour Board,[2] though the conditions of employment are determined by national and local agreements negotiated between employers and trade unions. There is no national agreement to the effect that dock workers should belong to an appropriate trade union, nor is such a condition laid down by the Order.

Before workers can obtain employment in the docks they must be accepted for registration as dockers by the local Dock Labour Boards, which are comprised of an equal number of representatives of workers and employers. With a given number of vacancies on the docks, trade unions and employers are each allowed to nominate half of the workers

[1] *Trade Union Documents*, by Milne-Bailey, p. 461.

[2] For full particulars of this scheme see Statutory Rules and Orders, 1947, No. 1189.

required. Obviously all of the trade-union nominations are trade unionists, and the employers normally agree that their nominations should join a union too. In London, Liverpool, Manchester, and Bristol local agreements require a nominee to produce a clearance certificate from his union stating that his contributions are fully paid up before he can be registered. Once this has been complied with the influence of the local Dock Labour Board ceases to operate as long as the man accepts the conditions laid down by the Order. If, however, a board dismisses a docker who has been expelled from his union, the man has a right of appeal to a specially constituted appeals tribunal. Even where local agreement requires membership of a trade union as a condition of employment, it is doubtful whether a docker expelled from his union could be legally dismissed from the docks. A test case has yet to be heard.

In the coal mines trade-union membership is a condition of employment in certain districts only. In January 1948 the National Union of Mineworkers asked the National Coal Board to make a national agreement compelling most mineworkers to be members of that union and providing for the deduction of union dues from their wages. The board refused to accede to this request for three reasons: (1) if a man were expelled from the union for a breach of union regulations the board would be obliged to dismiss him even though the board had no complaint against him; (2) all workers who did not want to join a trade union, but who were already in employment in the mines at the commencement of an agreement, would have to be dismissed regardless of their efficiency or length of service; (3) as, since nationalization of the mines, there is only one employer, workers would not merely be discharged from a particular job but from the industry, with little likelihood of being able to use the skill and experience they had acquired in the mines.[1] In the coalfields of Durham, north Wales, south Wales, and Warwickshire trade-union membership was already a condition of employment imposed by local agreement before the National Coal Board took over the industry. Mineworkers in north Staffordshire, south Staffordshire, Cannock Chase, and Shropshire were also required to join the appropriate union but were given the option of 'contracting out'. The board stated that it was prepared to continue with the existing agreements and also to encourage complete and orderly representation amongst mineworkers; furthermore, it was prepared to deduct trade-union dues from the wages of workmen provided a satisfactory procedure could be worked out.[1] This has since been done.

Workers who are almost indispensable in a highly mechanized production process and cannot easily be replaced by machines are in a strong

[1] *National Coal Board, Report and Accounts*, 1948, p. 48.

position to control the entry of workers into their trade and, therefore, to negotiate a 'closed shop' agreement. Even so, experience has shown that a trade union representing such workers is not strong in times of unemployment unless the possession of that union card is a positive criterion of a workman's skill. This factor is undoubtedly the most important and most permanent of all factors influencing employers in their attitude to compulsory trade unionism. Craft unions normally operate under these conditions. The only variable factor, so far as the unions are concerned, is the control of entry into their respective crafts, and this, by their regulations, they attempt to influence. In different industries they have experienced varying degrees of success. Some unions have practised 'patrimony', that is, restriction of entry to the sons of craftsmen or workers in a particular trade, though this has acted more as a form of nepotism than a regulative influence. Undoubtedly the most widely used and the most successful method of controlling the supply of labour into a trade has been the regulation of new entrants through apprenticeships.

To be successful as a regulative influence a period of apprenticeship must, in the words of Sidney and Beatrice Webb, be a 'period of educational servitude and formal testing of competency'; otherwise employers could upgrade semi-skilled or unskilled men to skilled jobs. In most of the building crafts the unions have been unable to control the entry of workers because of the comparative ease with which building labourers have been able to learn enough about a craft to be taken on by employers in the place of apprentice-skilled workers. Unions organizing bricklayers, painters, plasterers, and tilers have been particularly affected by this unregulated source of labour. And once a trade union has failed to limit the entry into its trade it very soon loses interest in educational standards.[1] In the engineering industry craft unions have, on the whole, likewise failed to regulate the flow of labour, but for different reasons. The rapid expansion of the engineering industry, especially that section of it using mass-production methods, and consequently the continuously rising demand for labour, has made nonsense of any attempt to restrict labour supplies. In addition, the wide use of mechanization has removed from a large range of engineering jobs the element of skill necessary to justify apprenticeships. In the motor-car and aircraft sections of the engineering industry it is common practice to upgrade semi-skilled men to skilled jobs, thus making it virtually impossible for unions to control labour supplies.[2] There are exceptions of course. In such crafts as the pattern-makers, sheet-metal workers, coppersmiths, and heating and domestic engineers

[1] *Industrial Democracy*, by S. and B. Webb, p. 478.
[2] *Productivity and Trade Unions*, by F. Zweig, p. 228.

it would be difficult for anyone to get a job without possessing a trade-union card.[1] There are several trades in the cotton industry, such as mule spinners, strippers and grinders, loom overlookers, tapesizers and cloth overlookers, organized by craft unions which can claim fully comprehensive trade-union organization. Some of these unions have negotiated 'closed shop' agreements and through the method of 'patrimony' virtually enforce a closed union too.[2]

The conditions that give rise to compulsory trade unionism can be seen in their totality in the printing industry. A number of operations in the industry are undertaken by highly skilled craftsmen who have been able to enforce rigid apprenticeship regulations and who have a long tradition of trade-union organization behind them. Employers know that a worker possessing the card of a printing craft union has, in most cases,[3] been through a period of rigorous craft training. This is an aid to the efficiency of printing establishments, for a process of selection of workers which involves trial and error can be a costly process in jobs where skill is required. Each craft union, in agreement with the employers, enforces its own apprenticeship regulations. For example, the Printing Machine Managers' Trade Society, which organizes machine-minders in London, in an agreement with the London Master Printers' Association has fixed the ratio of apprentices to journeymen and has stipulated that apprentices shall serve at least five years of actual craft training. The agreement[4] states: 'there shall be a periodic review of the labour supply situation with a view to reaching agreement on any adjustments which could deal with any surplus or shortage in labour supply'. If there is a continued shortage of skilled labour the parties to the agreement shall consider (i) revision of the apprenticeship quota, (ii) the introduction of an additional block of apprentices, or (iii) any other means of dealing with the situation. In a period of continued shortage of skilled labour the efforts of any union to restrict rigidly the supply of labour would undoubtedly be undone by employers, either through changes in production methods, or, where these are not possible, by violation of the agreements. In order, then, for trade unions to retain some control over the labour supply during periods of change it is essential that they should, by agreement, be consulted over any changes in practice to meet new situations. This is particularly important when unemployment occurs. The above-mentioned agreement

[1] Ibid., p. 225.
[2] Ibid., pp. 120–1.
[3] Controlled upgrading is permitted by craft unions. An example is given below.
[4] This agreement made on 30 May 1951, which also determined minimum wage-rates for specific operations, overtime, and notice of dismissal, was calculated to 'remain in force for a period of five years from 16th November, 1950, subject to adjustments as provided in circumstances specifically referred to in clauses in the Agreement and to the operation of the cost of living sliding scale'.

states that in the event of unemployment rising to 4 per cent. of the total journeymen working membership, consideration shall be given to (*a*) a reduction within reasonable limits of the number of hours in the normal working week, (*b*) the encouragement of retirement by journeymen over the age of 65, and (*c*) any other means of dealing with the situation. This agreement illustrates the mechanism of an effectively 'closed shop'.

Similar, though not so comprehensive, agreements exist outside of London. There has been virtually no dilution in the printing industry and there is no upgrading of semi-skilled men to skilled jobs in any section of the industry without union agreement. An arrangement exists in the provinces between the Typographical Association, a craft union, and the National Society of Operative Printers and Assistants whereby experienced printers' assistants who are members of the latter union, which organizes unskilled and semi-skilled workers, are given equal opportunity with the members of the Typographical Association for promotion to machine-minders after unemployed machine-minders available in the area have been absorbed. The upgraded workers must, however, transfer to the craft union in accordance with the terms of the agreement.[1]

Some Comparisons with the U.S.A.

A recent survey by the U.S. Bureau of Labour Statistics found that over 70 per cent. of the workers covered by labour agreements in the U.S.A. now have some form of 'union shop'.[2] In a more detailed study undertaken by the American National Industrial Conference Board it was found that out of 602 post-Korean War trade-union contracts only 196 did not contain any type of trade-union security provision. The full 'union shop' occurred in 207 of the contracts, maintenance of membership in 121,[3] and a modified 'union shop' in most of the remainder with trade-union security provisions. Three million workers were covered by the study, and of these only 314,000 had no form of security clause in their contracts. 'Union shop' agreements covered 465,000 workers in the coal-mines, 242,000 workers on the Eastern railways, and 100,000 workers employed by the Ford Motor Company. 'Check off' arrangements, whereby union dues are deducted from the worker's pay by the employer, were provided for in 435 of the contracts examined.[3] Recently concluded agreements have taken different forms. For instance, the agreements concluded between the railway unions and the Eastern and Western railway companies in

[1] Agreement (Rotary Department) between the Typographical Association and the National Society of Operative Printers and Assistants, as amended in June 1948.

[2] *Labour News from the U.S.*, 1 Aug. 1952, vol. ii, No. 53. U.S. Embassy.

[3] Maintenance of membership makes union membership in good standing a continuing requirement for holding the job; if the worker fails to maintain good standing, the employer must dismiss him (*Labour News from the U.S.*, 9 Apr. 1953, vol. iii, No. 8).

the U.S.A. require union membership in 60 days after reporting to work; whereas as a result of the 8-week national steel strike in the U.S.A. which ended in July 1952, steel workers must sign a union card at the commencement of employment, but they are given a 15-day period at the beginning and end of each contract during which time they can withdraw from membership of the union.[1]

The 'closed shop' is made illegal in the U.S.A. by the Taft–Hartley Act, 1947, though the same Act permits the 'union shop' except where State legislation rules to the contrary.[2] Section 8 (*a*) (3) of the Act states: 'Nothing in this Act, or in any other statute of the United States shall preclude an employer from making an agreement with a labour organisation . . . to require as a condition of employment membership therein on or after the thirtieth day following the beginning of such employment or the effective date of such agreement, whichever is the later.'[2] An agreement can only be concluded after the workers concerned have approved it through a secret ballot conducted by the National Labour Relations Board. An application for a ballot must be supported by at least 30 per cent. of the workers in an establishment. As a result of a ballot a union can be given exclusive bargaining rights for a minimum period of 12 months, after which time workers can apply for another ballot.[2]

What is the reason for the endemic nature of compulsory trade unionism in the U.S.A. as compared with Britain? Since the repeal of the 1927 Trade Disputes and Trade Unions Act in 1946 there has been no legal prohibition of compulsory trade unionism in any field of British industry; yet it is far from being a widespread practice. In an attempt to explain the difference an American writer once wrote: '. . . in the United States . . . the presence of great numbers of immigrants, the relatively ample opportunities for workers to rise out of the ranks of labour . . . have made the development of . . . a feeling of class solidarity impossible. Threatened . . . by those who were willing to accept low wages and a long working day because they considered their position as wage-earners only temporary, and . . . by new groups with lower standards of living, most American trade unions saw in the closed shop their only protection.'[3] He pointed out that the greater degree of homogeneity among European workers and the higher proportion of trade unionists resulting from this homogeneity have decreased the stress placed on 'closed shops' by European trade unionists. This is clearly a faulty analysis, for solidarity is in all cases a

[1] By the end of 1952 it had been estimated that approximately 5 per cent. of the new entrants to the industry were taking advantage of the provision to contract out of the union, the United Steel Workers (letter from Professor Harbison, University of Chicago).

[2] United States Department of Labour, Bulletin No. 123, August 1950.

[3] 'Closed Shop and Open Shop', by W. M. Leiserson, *Encyclopaedia of the Social Sciences*, 1930. Quoted in *Unions, Management and the Public*, edited by Bakke and Kerr, 1949.

prerequisite for negotiations for 'closed shop' agreements and it is where solidarity is greatest among workers that one finds the 'closed shop' principle most firmly established. In Britain, the dockers, miners, and printers possess a deep feeling of solidarity. The protracted and somewhat bitter negotiations between the United Steel Workers and the steel employers in the U.S.A. over compulsory trade unionism would never have taken place if the steel workers had lacked solidarity in any sense. Non-unionists can only be forced into a trade union if that union already organizes a high proportion of the workers in its special field and if employers agree to have their workers forced into a union.

Labour in Britain, because of its relative abundance, has tended to be cheap in relation to the cost of capital. The general effect of this has been to cause entrepreneurs to use extensive methods for increasing production rather than intensive methods through a greater use of machinery. Once methods are established which depend for efficient production on the availability of cheap labour, then employers have a vested interest in obtaining the highest possible degree of labour utilization. It follows that where this occurs the acceptance of a labour policy which removed from employers the right to choose and eliminate workers according to their efficiency would be entirely uneconomic for employers. In the U.S.A., however, the forces have operated in reverse. There, labour has been persistently expensive relative to capital with the consequent effect of a high degree of mechanization in industry. This, coupled with an extensive use of continuous-production methods, has placed employers in the U.S.A. in the position whereby, when faced with determined trade-union demands for compulsory trade unionism, they accept it as the lesser of two evils.

It is relatively simple to give reasons to explain the attitude of employers to compulsory trade unionism. Trade-union attitudes to it are much more abstruse. Under similar conditions American trade unions have always placed greater emphasis on achieving compulsory trade unionism than have British unions which have been more concerned with securing the right to negotiate for all workers rather than to organize all workers. British unions have insisted that if some workers prefer to stay outside of trade-union organizations they must, nevertheless, accept within the precincts of the factory the conditions agreed upon through the process of collective bargaining.[1] This demand has been accepted by employers and has eliminated, almost completely, individual wage bargaining in British industry. American trade unions have endeavoured to achieve this end by getting all workers into their organizations. Other explanations of the different trade-union attitudes towards compulsion

[1] *British Trade Unions*, by N. Barou, p. 95.

lie in the development of a tradition of voluntary allegiance in the British trade-union movement which, because of its relatively recent growth and dependence on the support of the legislation of the New Deal era, has not been developed by the American trade-union movement to the same extent; and in the different approaches of British and American trade-union leaders to institution building.

Conclusion

When the demand for compulsory trade unionism comes from rank-and-file members it is usually because they object to workers receiving trade-union benefits without contributing in some way to the maintenance of trade-union organizations. It is a question of equity. The motives of trade-union leaders are more complex. Often compulsory unionism agreements are regarded as 'institution building' devices, motivated by the desire to remove the problem of recruitment, to seek an ally in the enforcement of discipline in the union, or simply to acquire power. Compulsory membership of craft unions is a device to regulate the supplies of labour, and, therefore, to determine wage-rates. In any case, compulsion in trade unions raises more problems than it solves.

The attitude of employers will depend on what Professor Hicks has called the 'employers' concession curve' which is the relative costs of concession and resistance. Concession cost is the effect of compulsory trade unionism on the efficiency of labour in a business organization. If efficiency improves, as it can do when trade-union membership indicates skill and training, then no problem exists. It is a different matter when efficiency is adversely affected through a reduction in the managerial control of labour. Then an employer has to calculate the cost of resisting the demand for compulsory trade-union membership. This cost depends on the proportion of his workers supporting the demand, the action they, or the trade unions representing them, are likely to take to achieve it, and the effect of the proposed action on costs. If these costs are high, as they are likely to be if the production methods are highly mechanized, or if a vital public service is provided, then the employer may accept compulsory trade unionism; if they are not high, as, for example, in the building industry, in relation to the effect on costs of losing the effective control of labour, then the employer would probably resist the demand for the compulsory enrolment of his employees into any trade union.

OXFORD UNIVERSITY DELEGACY FOR EXTRA-MURAL STUDIES.

AN EXPERIMENT IN COMMODITY CONTROL—THE INTERNATIONAL WHEAT AGREEMENT, 1949–53

By C. D. HARBURY

I

THE fifth International Wheat Conference which met in London in March 1947 introduced a new scheme for stabilization of the price of a primary product which involved neither buffer stocks nor output controls. After two frustrated efforts and some alteration, that scheme came to fruition in the shape of the International Wheat Agreement in August 1949, and has since run to completion in July 1953. The overwhelming need for action in this field and the unsatisfactory nature of the remedies which had been adopted in the years between the wars were responsible for the considerable attention which the introduction of the agreement attracted. With the close of its 4-year period of life it now becomes possible to examine its effects, and this article is an attempt to measure the extent to which the experiment may be said to have succeeded.

The details of the agreement have been well documented,[1] but the main features may be briefly recalled. In principle, the agreement took the form of a multilateral contract between importers and exporters, such that the former promised to buy, and the latter to sell, certain fixed quantities of wheat to each other in each year of the agreement's life. Maximum and minimum prices were also decided upon,[2] and the agreement so worded that the guaranteed purchases and sales would operate only at the price limits. Thus while the free play of market forces could operate within these limits, as soon as the price rose above or fell below them, either the sellers would be called upon to sell at the maximum or the purchasers to buy at the minimum price. The price of wheat covered by the agreement, therefore, must be contained within the limits set. For wheat traded outside the agreement, on the other hand—and it was an important feature that the agreement should not cover all of world trade in wheat (even all of that of the parties to the agreement)—price was free to fluctuate with changes in supply and demand. This dual price feature, together with the guaranteed quantities, gave the whole scheme something of the elegance of the theoretical economist's model and the attractiveness of a compromise between the twin aims of commodity policy—stability

[1] See the agreement itself, Cmd. 7680; H. Tyszynski, 'Economics of the Wheat Agreement', *Economica*, Feb. 1949; Frank H. Golay, 'The International Wheat Agreement of 1949', *Quarterly Journal of Economics*, Aug. 1950.

[2] The maximum remained constant, while the minimum was to be progressively reduced in each of the 4 years from $1.50 to $1.20.

simultaneously with the adjustment of supply to changes in demand. The guaranteed purchases and sales were to provide sufficient stability[1] for both producers in income and consumers in expenditure, while the barometer of the price mechanism was to operate in the partial free market and register the necessary adjustments.

Any appraisal of the agreement therefore must begin by investigating, on the one hand, the stability, and on the other the degree to which supply has kept pace with changes in demand. One difficulty must be faced immediately. The wheat agreement was an agreement between governments, and, more important, between governments on the export side, which happened to operate very positive agricultural policies, and on the import side, which happened to exercise import monopolies for wheat.[2] In these circumstances, such stabilizing effects as there may be must operate on the foreign-exchange receipts of the exporting countries and the expenditures of the importing countries, rather than directly on the consumers and producers of wheat. This is of particular concern on the export side; but, in so far as the producers' governments operate stabilization programmes for producers (as do all the exporters), whether or not they operate export monopolies (as do Canada, Australia, and France), and whether or not they pass on more or less than the full benefit of the agreement, the observer has no choice but to judge the efficacy of the agreement in this connexion by the effects upon governments. While this is no great loss on the count of stability, where foreign-exchange reserves are even more important than farmers' incomes, it is less true on the count of adjustment. More will be said in this connexion when dealing with production changes, but at this stage it is important to be aware of the limitations to an ideal appraisal which are, as usual, set by the facts of the real world.

The arrangement of the discussion is as follows. Section II deals with the experience of stability, and section III with adjustment. Section IV is concerned briefly with certain general matters regarding administration, and in the last section, V, an overall appraisal is attempted.

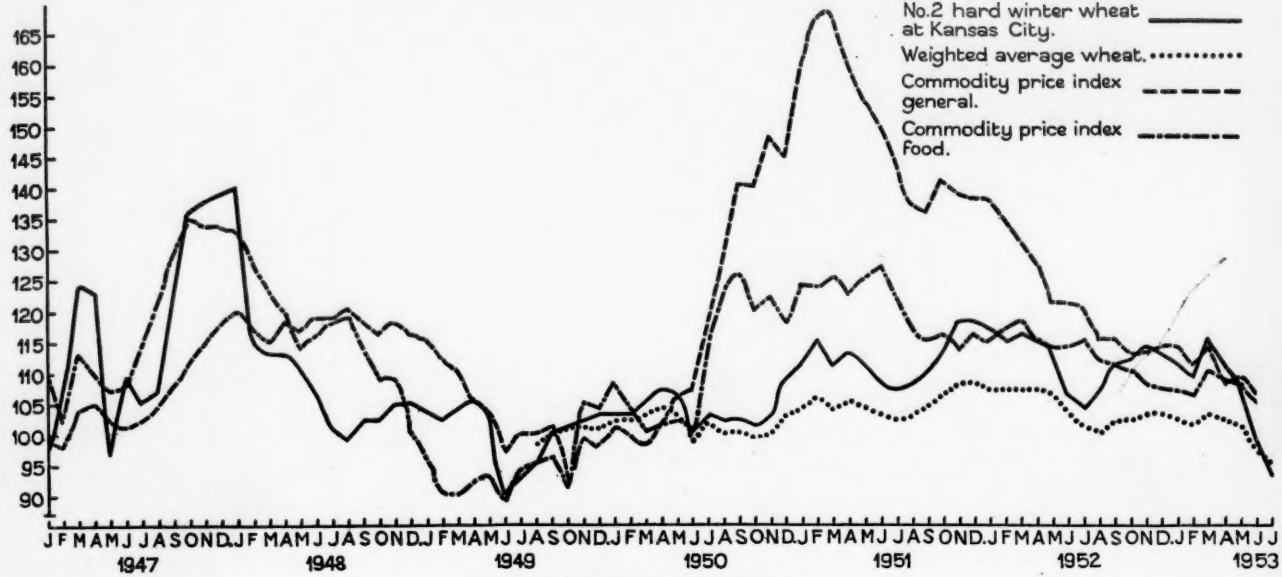
II

The measurement of the exact amount of stability engendered by the wheat agreement cannot, of course, be calculated, for we can never have

[1] In the ill-fated 1947 and 1948 agreements even greater stability was provided for by guarantees concerning stocks, similar to those concerning sales and purchases, by parties on both sides. Although this provision was left out of the 1949 agreement, it had not been this so much as the actual fixing of the prices which had been the stumbling-block.

[2] Over 80 per cent. of the imports in 1949–50 of countries parties to the agreement were purchased by State trading monopolies. See Commodity Policy Studies, No. 1: *A Reconsideration of the Economics of the International Wheat Agreement*, F.A.O., Sept. 1952, Appendix D, Table III. B.

Indices of Monthly Commodity Prices. (1949 = 100)


No.2 hard winter wheat at Kansas City.
Weighted average wheat.
Commodity price index general.
Commodity price index Food.
165
160
155
150
145
140
135
130
125
120
115
110
105
100
95
90
J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J
1947
1948
1949
1950
1951
1952
1953

Sources and Notes—See Appendix.

more than a rough idea of the hypothetical price of wheat in the absence of the agreement. This, however, is the familiar difficulty which arises from the impossibility of controlled experiment in the social sciences, and which is usually circumscribed by the process of successive approximations. Here we may approach the ideal by following, first, the earlier behaviour of the price of wheat, and second, the prices of other commodities.

But before embarking on such a study we must decide which wheat price shall be taken as representative of the stability produced by the wheat agreement. Clearly the prefixed price limits are no use, since they do not cover all the trade in wheat, and would therefore give the agreement too favourable a showing; and, on converse reasoning, the free market price is no better. The only sensible price to take for the purpose of measuring stability of expenditure on, or revenue from, sales of wheat, as has been pointed out, is the weighted average price on agreement and non-agreement transactions.[1] For the extent of fluctuation in expenditure on wheat[2] depends on two factors—fluctuations in the free market price, and the proportion of guaranteed quantities to total trade in wheat.

If we take the ratio of agreement to total trade in wheat as the weight to be attached to the fixed wheat agreement price,[3] and the complement of it as the weight for the free market price, we can calculate a weighted average price of wheat which measures the stability of total expenditure with the agreement in force. Table I shows the amount of world trade in wheat inside and outside the agreement, for the years 1949–53, and provides the weights for the calculation of this average price, which is shown in the form of an index number based on 1949, in the chart opposite, and is plotted alongside a similar index of the representative free market price.

Inter-temporal comparisons are notoriously untrustworthy, since far too many variables are usually changing at the same time, but by comparison with the fluctuations in the price of wheat from 1947 to 1949 the stability of the average wheat price after the introduction of the agreement is certainly striking. This relative stability is further emphasized in Table II, which brings together the average year-to-year fluctuations in the market price of wheat for various periods in this century. The figure of 2·7 per cent. average price fluctuation during the course of the agreement compares very favourably with all periods since 1901, even with that previously most stable first decade of the century, when the average fluctuation was 5·7 per cent.

[1] See C. D. Harbury, 'Commodity Agreements and Price Fluctuations', *Economic Journal*, Sept. 1951.

[2] Since total world expenditure on and receipts from sales of wheat amount to the same thing, the discussion is in future conducted only in terms of the former. This device for ease of expression should not cause one to overlook this fundamental identity.

[3] The wheat traded under the agreement was all sold at the maximum price of $1.80.

TABLE I

World Trade in Wheat and Wheat Flour, 1949–53

Million metric tons (in wheat equivalent)

Crop year	*Recorded under wheat agreement*	*Total world exports*	*Ratio* $\frac{\text{Agreement}}{\text{Total trade}}$ $\frac{(2)}{(3)} \times 100$
(1)	(2)	(3)	(4)
1949–50	11·8	22·6	52:48
50–1	14·5	25·2	58:42
51–2	15·6	28·0	56:44
52–3	15·0*	23·1†	65:35

Notes: * To 26 June 1953.
† Estimate from *The Economist*, 25 July 1953 (p. 275).

Sources: Annual reports, International Wheat Council; Report No. 50, *Cumulative Totals of Sales and Purchases for Crop Year 1952–3*, International Wheat Council; *Yearbook of Food and Agricultural Statistics*, F.A.O., United Nations; *Grain Exports by Source and Destination*, F.A.O., United Nations.

TABLE II

Year-to-year Fluctuations in Market Prices of Primary Products: 1901–53*

Commodity	*1901–50*	*1901–13*	*1914–19*	*1920–39*	*1940–5*	*1946–50*	*1949–53*	
							Weighted† average	*Free market*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Wheat .	11·9	5·7	14·0	15·3	12·9	9·8	2·7	4·3
Foodstuffs	13·9	8·0	15·7	15·3	14·4	13·8	..	..
All . .	14·0	10·3	17·9	14·6	11·0	16·0	..	..

Notes: * Average fluctuation per year.
† Calculated from figures given in the Appendix.

Source: *Instability of Export Markets in Under-developed Countries*, United Nations Department of Economic Affairs, Sept. 1952.

By themselves, of course, these comparisons can reveal little, for it might well be that there were other special factors, in addition to the international wheat agreement, making for stability of wheat prices from 1949 to 1953. That this was not in fact the case is widely recognized, for 1950–1 was a year of almost unprecedented fluctuation in primary product prices. The general movement of commodity prices over this period is well-known, and is brought out in the accompanying chart, where general and food indices of monthly primary product prices have also been plotted alongside the movements in the average price of wheat. Before the international wheat agreement, it will be seen, the price of wheat moved roughly parallel to the price index of foodstuffs, and actually fluctuated

more violently than either that or the index of general commodity prices. In the course of the 1950–1 commodity boom, however, when the general index mounted nearly 70 points, and the food index 27 points, the average price of wheat rose no more than 8 points—by contrast, a model of great stability.

TABLE III

Fluctuations in Monthly Food Prices, 1949–51

(1949 = 100)

Commodity	*Grade and market*	*Percentage increase for peak month 1951*
Barley*	U.S. No. 3 cash price Minneapolis, U.S.A.	123
Barley	No. 1 feed barley in store Fort William/Port Arthur, Canada	125
Rye*	U.S. No. 2 cash price Minneapolis, U.S.A.	129
Rye	No. 2 C.W. price to producers for domestic use and for export, basis in store Fort William/Port Arthur, Canada	161
Maize	No. 3 Yellow cash price at Chicago, U.S.A.	142
Rice	Average price to farmers, U.S.A.	142
Sugar	96-degree basis f.o.b. Cuba	176
Tea	Medium Assam Broken Pekoe for export, India	110
Coffee	Santos No. 4 at Santos, Brazil	179
Cocoa	Superior Type at Bahia, Brazil	252
Wheat	No. 2 Hard Winter at Kansas City, U.S.A.	118
Wheat	Weighted average price of agreement and non-agreement trade	108

Notes: * Base, average of 1948–9 and 1949–50 averages.

Sources: Monthly Bulletin of Food and Agricultural Statistics, F.A.O.; *Survey of Current Business*, U.S. Department of Commerce; *Review of International Commodity Problems 1951*, United Nations.

Further support is given to this picture by comparison with the prices of other important food products in international trade. Table III gives details of the percentage increases on 1949 of the peak prices in important world markets in 1951 for rye, barley, maize, rice, cocoa, coffee, tea, and sugar. In every case the price of wheat has been more stable, its peak rise of 8 per cent. comparing with an average of 29 per cent. for cereals, over 70 per cent. for coffee, sugar, and cocoa, and 10 per cent. for tea. The weight of evidence is certainly overwhelming. In spite of fears in some quarters that the proportion of world trade covered by the agreement might be insufficient to ensure stability, it seems safe to conclude that, in at least this objective, the experiment of the wheat agreement has been successful.[1]

[1] Joseph Lerner, in a very neat theoretical construction which has come to my notice since this was written, has furnished a formal argument which reinforces this conclusion concerning the stability of the average price of wheat. He shows that, given the most general rules concerning the shape and distribution of indifference curves, the destabilizing income effect generated by the wheat agreement cannot outweigh the stabilizing effect of the fixed price on agreement sales. (*Quarterly Journal of Economics*, May 1953.)

It does not, of course, follow that this general picture of world expenditure on wheat applies in detail to each and every wheat-trading country. There are two reasons for this. In the first place, individual countries signing the agreement covered different proportions of their total trade in wheat by guaranteed quantities, and it is only for the 'average' country, or for any countries whose trade was divided between agreement and non-agreement wheat in the same proportions as total world trade, that this is true. In the second place, it has been assumed that all trade outside the agreement took place at the price of No. 2 Hard Winter at Kansas City, or at rates which fluctuated parallel to it. While this is not an unfair approximation, it is certainly not univerally true, particularly of the trade of countries which were not parties to the agreement. In so far as the two most important of such exporters (Argentina and the Soviet Union) have State trading monopolies which conducted long-term contracts in wheat, there is good reason to believe that any differences would be such as to make the degree of stability given in the picture drawn above, greater rather than less.

III

It is a common feature of most types of commodity schemes that the more efficient they are at stabilizing income or expenditure, the less efficient they are at causing adjustment of supply to changes in demand. This follows naturally from the suppression of the free working of the price mechanism, and is certainly typical of buffer stocks.

One of the greatest attractions of the wheat agreement at the time of its introduction, however, was that this was held to be less likely. For a portion of the market was still subject to the free play of supply and demand, and the price mechanism would be allowed to work there and provide the necessary incentive for adjustment. Indeed, it was even argued that the introduction of the wheat agreement, with guaranteed quotas, would reduce the income effect of a change in price on consumers, and would thereby reduce the elasticity of demand in the free market—so increasing the intensity of price fluctuation there, and acting as an even greater stimulus to alter production with shifts in demand.[1] In so far as this may be the case, the agreement may be given a good record on the adjustment count, since the magnified price stimulus would act both to

[1] See H. G. Johnson, 'The Destabilising Effect of International Commodity Agreements on the Prices of Primary Products', *Economic Journal*, Sept. 1950; F. H. Golay, op. cit.; C. D. Harbury, op. cit. It has been also argued that the introduction of the agreement would reduce the income effect of a price change on the producers of wheat, and thereby increase the elasticity of supply and offset the decrease in elasticity of demand. Since the agreement itself has had little direct effect on the actual producers of wheat, it seems reasonable to disregard the changes in elasticity of supply here.

decrease production generally with a fall in price, and, other things equal,[1] to put the least efficient farms out of business first.

Since, as has been explained earlier, most producing countries adopt positive agricultural policies, the best test of the efficiency of the wheat agreement in this connexion is not the progress or decline of production but the behaviour of price in the free market. For even though producers' governments may not pass on to farmers the full effects of changes in the free market price, it is probably true to say that their long-term policy with regard to wheat production—be it ever so mandatory—must at least be influenced by the changes in this price.[2]

Once again the problem arises of how to compare the price of wheat outside the agreement with what the price would have been in its absence, and again we are reduced to the same two alternatives. Referring once more to the accompanying chart, the free market price, as represented by No. 2 Hard Winter at Kansas City, can now be compared, first with the same price before the agreement came into force, and second, with the prices of other commodities. It will be seen that, although naturally the free market price fluctuated from 1949 to 1953 considerably more than the average price which was previously used as a measure of stability, it fluctuated very much less after June 1949 than before that date. Again, compared with indices of general and food commodity prices and with the prices of the same individual foodstuffs, the free market price of wheat was more stable. In the case of only one commodity—tea—where a producers' restriction scheme was in force, did it fluctuate more. Finally, even further confirmation of relative stability is given by the comparisons of year-to-year fluctuations in wheat and other commodity prices since 1900 in Table II. Once again the evidence is overwhelming that, in the face of the commodity boom of 1950–1, the wheat agreement resulted in a more stable price even for wheat sold outside the agreement.

While the wheat agreement does not merit a favourable comment on these grounds, it is interesting to consider why it should have acted in this way. There are two possible explanations. In the first place, it was previously argued that the reduction of the income effect would decrease the elasticity of demand. In the case of inferior goods, however, a reduction of income effects operates so as to *increase* the elasticity of demand. Thus with a given increase or decrease in demand price rises or falls less as a result of the agreement. While it is not impossible that wheat is on

[1] This *ceteris paribus* assumption is not intended to blind the eye to possible perverse effects of price changes on primary producers, since the agreement works directly on governments whose reactions are more likely to be normal.

[2] In this connexion it has been recommended that participating governments operate dual pricing policies with farmers in the home market, in order to pass on the incentive in full, rather than rendering the same sum of money down to an average price. See F.A.O., op. cit., para 26.

balance an inferior good in world consumption, it seems unlikely that the income effect can be strong enough to account for the considerable influence on price which has been brought out here.[1]

The second, more plausible, reason lies in the effect of the agreement itself on changes in demand rather than on the elasticity of the demand curve. The 1950–1 commodity boom was predominantly a speculative occasion caused by panic-stricken governmental stockpiling following the outbreak of war in Korea and the fear of short supplies. It seems eminently reasonable to suppose that the existence of the guaranteed sales under the wheat agreement, even though they represented only just over half of world trade in wheat, were sufficient to allay these fears as far as wheat was concerned. The speculative demand was thus channelled into other commodities, and even the free market price of wheat remained stable by contrast with almost all other foodstuffs and raw materials.

IV

History of past attempts at international commodity control, let alone control by an association of the opposing interests of consumers and producers, would hardly give great confidence that the wheat agreement would run smoothly. Nevertheless, not only did the agreement not break down, but positive praise was forthcoming for the efficiency of its administration on more than one occasion.[2] Accordingly some brief attention should be paid to this very remarkable performance by considering the more important day-to-day problems which arose and the ways in which they were solved.[3]

Articles XIII–XIX of the agreement laid down rules for its administration which set up an executive council with power to appoint sub-committees where necessary. One of the most important of these was entrusted with the task of calculating the appropriate price equivalents for all varieties and grades of wheat for individual transactions, taking into account variations in exchange rates between currencies. This intricate task which involved a great number of pitfalls was performed, in almost every case, to considerable satisfaction.

Another important function of the council concerned the recording of all transactions under the agreement between the parties. Evidence of the satisfactory completion of this duty is presented by the challenges of

[1] While the income effect on the demand side is usually considered unimportant in economic theory, the existence of State trading and the preoccupation with foreign-exchange reserves may well increase it.

[2] See Gordon P. Boals, 'International Wheat Agreement—its First Year', and 'International Wheat Agreement—its Second Year', *Foreign Agriculture*, Oct. 1950 and Oct. 1951.

[3] This section draws heavily on the annual reports of the International Wheat Council, whence, except where otherwise stated, all factual material derives.

transactions which under the terms of the agreement were liable to be made. During the first two years these were negligible, and even in the third, although the council was somewhat disturbed at the increase, they amounted to no more than 3 per cent. of the total recorded. Table IV gives details of these transactions, and it will be seen that, with the exception of the first year, sales and purchases amounted to over 94 per cent. of the total guaranteed quantities under the agreement. In the case of 1949–50 the shortfall was largely due to new countries acceding at a date after the coming into effect of the agreement, and if allowance is made for this, the percentage fulfilment is raised to 93 per cent.[1] Other shortfalls were due to the failure on the part of importers (particularly Italy and Brazil) to take up their quotas,[2] for with the free market price well above the maximum, there could be no reason for exporters to demand purchase at the minimum. Requests for the assistance of the council under Article V by importers finding difficulty in purchasing their quotas were very rare and were satisfactorily arranged.

TABLE IV

International Wheat Agreement: Guaranteed Quantities and Transactions Recorded 1949–53

Crop year	*Guaranteed quantities*	*Transactions recorded*	*Percentage fulfilment* 3/4 × 100
(1)	(2)	(3)	(4)
1949–50	14,291	11,760	82%
1950–1	15,310	14,451	94%
1951–2	15,590	15,573	100%
1952–3	15,810	15,032*	95%

Note: * Year to 26 June only.

Sources: International Wheat Council, annual reports and Report No. 50, *Cumulative Totals of Sales and Purchases for Crop Year 1952–3.*

One of the most remarkable achievements, however, came about in 1951–2 when one of the exporters, Australia, indicated that she was likely to be faced with a short crop and might not be able to supply her full quota. Although the agreement made some general provisions for such an eventuality, the necessary adjustments were finally agreed upon by voluntary relinquishment of quotas by certain importing and exporting countries.

The only significant difficulty which arose, and which was not finally resolved by the council, concerned the carrying charges which Canada

[1] See Boals, op. cit., Oct. 1950.

[2] These were due principally either to over-estimation of requirements or to partial satisfaction in bilateral deals with an exporter not party to the agreement.

imposed in 1951–2 in addition to the agreement price, and which the United Kingdom opposed as being on technical grounds a breach of the agreement. The parties divided on the matter according to their interests as importers or exporters, and although 'there was every indication on both sides . . . of a desire that the question should be reasonably and amicably settled and that a prolonged deadlock should be avoided',[1] the disagreement was consequently never resolved. As the price plus carrying charge remained below the free market price, however, importers had no alternative but to accept the charges imposed.

A number of other matters were successfully dealt with, eight new countries were admitted to the agreement, and exporters agreed to a 10 per cent. increase in their guaranteed sales to cover these and other requests for increases from existing members—and with a total budget of between £30,000 and £40,000 per year, the administrative cost cannot be counted high.

V

It remains to consider the operation of the agreement in a somewhat broader setting. To a considerable extent this is possible after 4 years' experience, but it should not be forgotten that these particular years, in so far as they fit into any cycle at all, were boom years, whereas the efficacy of any contra-cyclical device can only finally be judged on its results during both upswings and downswings. Thus any statements or conclusions which follow, which refer to the latter stage of the cycle, can be no more than conjectures based upon experience in the opposite phase.

Of the number of criteria by which the wheat agreement may be judged, two are outstanding. The first—the measure of its success in terms of its own aims—may be disposed of quickly, for it formed the substance of sections II and III above.

Summarily it may be said that, on the side of stability of foreign-exchange receipts or expenditures, the agreement was undoubtedly successful. On its twin aim of adjustment, on the other hand, the record is somewhat less satisfactory, and while some incentive to adjust production was registered in the free market, this was very probably less than would have obtained in the absence of the agreement. The anticipation that the agreement might result in even greater price fluctuation in the free market has proved unjustified, and to that extent the theoretical argument in favour of a multilateral contract type of agreement loses force.

The second method of judging the agreement turns not so much on its internal consistency as on the alternatives which are available. In an

[1] International Wheat Council, *Annual Report, 1950–1*, p. 23.

agreement of this nature between conflicting interests, however, it is necessary to examine separately the effects on the two groups of parties to it—importers and exporters. In the case of the former this does not appear difficult. On the face of it, importers secured as a result of the agreement sufficient supplies of wheat for their requirements at an average cost to them of $202·5 per bushel,[1] whereas in the absence of the agreement they would have been almost certainly faced with a considerably higher figure.

Turning to the export side, it is therefore perhaps surprising to find general satisfaction with the results of the agreement. A typical attitude is represented by the statement of the U.S. Secretary of Agriculture after a visit to Europe in 1951 that 'local wheat prices in most countries I visited were below those in the United States. Were it not for the International Wheat Agreement, imported wheat would exceed the cost of local wheat and how long do you suppose we could expect to export our own surplus under those conditions?'[2] In Canada, too, the same attitude is found, as evidenced by the comment 'The International Wheat Agreement of 1949 . . . represents a fulfilment of an ideal which has been cherished by the wheat growers of exporting countries for over twenty years'.[3]

This general appearance of satisfaction with the working of the agreement in exporting countries, however, is in a sense false. It could more precisely be reckoned a satisfaction with the wheat situation itself, a state of affairs which was by no means due entirely to the international wheat agreement, but was rather the result of a partly fortuitous combination of circumstances including the tense state of international relations making for high stocks, intensified East–West trade difficulties, scarcity of rice in Asia, and, of dominant importance, the foreign aid programme of the United States.[4] While the relative importance of the agreement in the favourable course of events may not have been fully realized by all exporters, there is considerable evidence that they were aware of these

[1] The overall weighted average from 1949 to 1953.

[2] Quoted in Boals, op. cit., Oct. 1951. Another example of this U.S. attitude is the heading 'I.W.A. DISCOURAGES UNECONOMIC WHEAT ACREAGE EXPANSION' in a pamphlet intended for domestic consumption—*How is the Wheat Agreement Working?* Agricultural Information Bulletin, No. 74, Office of Foreign Agricultural Relations, Jan. 1952.

[3] D. A. MacGibbon, *The Canadian Grain Trade*, University of Toronto Press, 1952, p. 222.

[4] '. . . of the average exports of wheat from the United States of 411 million bushels annually from 1945–46 to 1951–52, an average of 229 (i.e. 55 per cent.) million bushels was financed by government funds'. Leslie A. Wheeler, Commodity Memoranda, No. 2, *Wheat*, International Association of Agricultural Producers. These funds, furthermore, though largely, were not reserved solely for U.S. wheat. For example, in 1949 the E.C.A. allocated $175 million Marshall Aid funds for the purchase of Canadian wheat by the United Kingdom. See *The Economist*, 24 Sept. 1949, p. 689.

other factors,[1] and also of the (albeit temporary) loss they probably suffered. Consequently, in the short experience of four boom years at least, there are strong grounds for concluding that of the two sides it was the importers who had the greater cause to be content with the agreement.

If this argument is correct, however, the experience of 1949 to 1953 will not explain the breakdown of the discussions for the renewal of the agreement, which came rather from the importers' side—or, more precisely, from the opposition of the United Kingdom, whose quota covered approximately one-third of the guaranteed purchases under the 1949 agreement. In fact, of course, the United Kingdom attitude is explicable in terms of the conditions of renewal, rather than of the experience of the agreement in action. This is not the place to discuss the merits of the United Kingdom's argument, but the principle involved is important. It was not so much dissatisfaction with the stability aspects as doubts concerning the mechanism of adjustment that lay at the root of her objections. For, given the existence of very positive output policies in the exporting countries, it has been contended, the free market price of wheat might remain permanently above the maximum without expansion in production taking place.[2] While this argument implies the open or tacit collaboration of the exporting countries, the existence of which cannot be investigated here, there seems to be no sound reason why the further operation of the wheat agreement should be necessary for its continuance.

Thus it seems that, while importers have not been dissatisfied with the effect of the agreement on stability in their expenditure on wheat in a period of rising prices, they are less pleased with the prospect of their maintenance by producers' controls when other prices are expected to fall. Exporters, it might fairly be said on the other hand, accepted, by and large, the short-term loss they suffered, and regard the agreement as an orderly means of marketing their output. The possibility that the objections of the most important importer may jeopardize the agreement altogether leads to a consideration of the alternatives. Excluding the rather remote likelihood of a general stabilization scheme for all primary products—on the lines of a commodity reserve currency or an international commodity clearing-house—there are, in essence, only two possibilities.

The first is to leave everything to the free play of the market. While this is a policy that is very appealing to importers at a time when world

[1] See Robert L. Gastineau, 'U.S. Wheat and Flour in World Markets', *Foreign Agriculture*, Jan. 1952.

[2] In fact, of course, further difficulty was met in the discussions on the 1953 agreement in the shape of proposals for new maximum and minimum prices. The difference between the best offer of the United Kingdom and that of the exporters, however, was only 5 cents. In spite of the fact that even this might have meant an increase of some $9–10 million in the annual cost of Britain's imports, it is difficult to believe that this was the only, or the most important, reason for her remaining outside the agreement.

prices are expected to fall, to embrace it is to herald catastrophe. There is no need to elaborate time-worn arguments to support this. The fluctuations in the price of wheat alone in the 14 years from 1920 to 1934, when it fell some 500 per cent. (from 115 to 23 shillings per quarter[1]), is sufficient reminder. In a downswing of the cycle, moreover, importers would, as in the 1930's, experience the repercussions which a generation of economists brought up on the theory of the foreign trade multiplier are hardly likely to ignore.

The second alternative is an international buffer-stock scheme in wheat. The pros and cons of such a scheme compared with a multilateral contract like the wheat agreement have been argued before.[2] Generally, however, in so far as buffer stocks provide no incentive for the adjustment of supply to demand, and do not stabilize incomes so well in times of short crops, they get rather the worse of the comparison.[3] Nevertheless, there are some grounds on which a buffer stock may be preferable: not the least of these is the fact that it is only feasible to initiate such a scheme in a period of surplus. Primarily on this account, an international buffer-stock agency has never been set in motion for any commodity, and no practical test of its theoretical merits has been possible. For this reason, in addition to the fact that in the event of the complete collapse of the wheat agreement it would be at least a good second best, such an experiment would seem very worth while.

It is almost certainly true that, given their present positive agricultural policies, exporters favour the wheat agreement, but would operate those policies in any event. Faced with this situation it is difficult to argue that the interest of both importers and exporters lies other than in the acceptance of the agreement by the importers—exchanging their micro-economic losses for macro-economic gains, and relying on international organizations and other measures (and pressures) to supplement the production stimulus afforded by the free market under the wheat agreement.

But the macro-economic arguments for commodity schemes are only

[1] See P. de Hevesey, *World Wheat Planning and Economic Planning in general* (London, 1940), p. 157.

[2] See H. Tyszynski 'A Note on International Commodity Problems', *Economica* (New Series), Nov. 1950.

[3] An exception to this is where the buffer-stock agency's price limits are adjusted automatically with changes in supply and demand. A somewhat similar proposal for built-in flexibility of prices was put forward by the F.A.O. (op. cit.). The idea was extensively discussed by a sub-committee of the Wheat Council with a view to incorporating it in the renewed wheat agreement of 1953, but the technical difficulties in finding suitable criteria for such automatic adjustment seemed to be the cause of its being discarded. (See *Intergovernmental Policies and Arrangements for Agriculture*, F.A.O., p. 23, where this is suggested.) A further disadvantage of such a mechanism, in any case, is that adjustments, particularly downwards, would occur automatically, precisely at the time when the world interest lies in keeping prices up in order to prevent cumulative repercussions.

valid if the coverage is wide, and an international wheat agreement or buffer stock in wheat is not enough. Agreements are needed for other primary products. The sterling area in particular stands to gain immeasurably from such schemes, especially in rubber, tin, and wool. Action in the separate fields should not be, and indeed is unlikely even now to be, unconnected. While buffer-stock arrangements should not be ruled out where suitable for particular commodities, the experience of the international wheat agreement, as this article has tried to bring out, has been such as to encourage its extension for other primary products.

UNIVERSITY COLLEGE OF WALES,
ABERYSTWYTH.

STATISTICAL APPENDIX

Indices of Monthly Commodity Prices 1947–53

(1949 = 100)*

Date (1)	Wheat: No. 2 Hard Winter at Kansas City (2)	Wheat: Weighted† average (3)	Commodity price indices‡: General (4)	Commodity price indices‡: Food (5)	Date (1)	Wheat: No. 2 Hard Winter at Kansas City (2)	Wheat: Weighted† average (3)	Commodity price indices‡: General (4)	Commodity price indices‡: Food (5)
1947 J.	97	..	99	110	1950 J.	103	102	101	108
F.	108	..	98	102	F.	103	102	99	104
M.	124	..	104	113	M.	105	103	98	100
A.	123	..	105	110	A.	107	104	101	101
M.	97	..	102	107	M.	106	103	106	102
J.	109	..	101	108	J.	100	99	107	100
J.	105	..	102	114	J.	103	102	118	115
A.	107	..	105	121	A.	102	100	127	123
S.	122	..	108	129	S.	102	100	140	126
O.	136	..	112	135	O.	101	99	140	120
N.	138	..	115	134	N.	103	100	148	122
D.	139	..	118	134	D.	109	103	145	118
1948 J.	140	..	120	133	1951 J.	111	104	158	124
F.	115	..	117	125	F.	115	106	167	124
M.	113	..	115	121	M.	111	104	168	125
A.	113	..	118	120	A.	113	105	160	122
M.	111	..	117	114	M.	111	104	154	125
J.	106	..	119	116	J.	108	103	150	127
J.	101	..	119	118	J.	107	102	144	122
A.	99	..	120	119	A.	108	103	137	117
S.	102	..	118	114	S.	110	104	136	115
O.	102	..	116	109	O.	113	106	141	116
N.	105	..	118	109	N.	118	108	139	113
D.	105	..	116	100	D.	118	108	138	116
1949 J.	104	..	115	96	1952 J.	117	107	138	115
F.	102	..	112	90	F.	115	107	135	117
M.	104	..	110	90	M.	116	107	131	118
A.	105	..	105	92	A.	115	107	128	115
M.	103	..	103	93	M.	113	106	121	114
J.	90	..	97	89	J.	106	103	121	114
J.	93	..	100	94	J.	104	101	120	115
A.	95	98	100	95	A.	107	100	115	112
S.	100	100	101	96	S.	111	102	115	111
O.	101	101	92	91	O.	112	102	113	110
N.	102	101	99	105	N.	114	103	113	108
D.	103	101	98	104	D.	113	103	114	107
					1953 J.	111	102	114	107
					F.	109	101	111	106
					M.	115	103	114	110
					A.	111	102	108	108
					M.	108	101	109	108
					J.	99	97	107	105
					J.	93§	95§	108	110

Notes: * Col. 2 is based on average for 1949, col. 3 for Aug.–Dec. 1949, and cols, 4 and 5 for July 1949–June 1950.

† The average of No. 2 Hard Winter at Kansas City and the maximum price under the International Wheat Agreement, weighted by the proportion of total world trade inside and outside the agreement (given in Table I, above).

‡ *The Economist* Commodity Price Indicator (dollar basis).

§ 1 July.

Sources: Survey of Current Business, United States Dept. of Commerce; *George Broomhalls Corn Trade News*; *Records and Statistics.*

AN ALTERNATIVE PROOF OF A THEOREM IN WELFARE ECONOMICS

By CHARLES KENNEDY

ONE of the better-known propositions in welfare economics is that the aggregative index-number criterion

$$\sum P_2 Q_2 > \sum P_2 Q_1$$

entails the Scitovsky criterion, i.e. that no reallocation of the quantities in the first situation could make every individual as well off as he is in the second situation. In his justly celebrated 1950 article in *Oxford Economic Papers*, Professor Samuelson drew attention to the fact that this proposition had not been rigorously proved; and he himself supplied the missing step in the proof with the help of an Edgeworth–Bowley box diagram.[1] I offer here an alternative proof, which seems to me to be both more simple and more general, in that it is not restricted to the two-person two-good case. In what follows I shall adopt Dr. Little's useful convention of writing capital P's and Q's when dealing with the group and small p's and q's when dealing with the individual.

It is now generally accepted that, for a consistent individual,

$$\sum p_2 q_2 > \sum p_2 q_1$$

is a *sufficient* condition for him to be better off in situation 2. It clearly follows that

$$\sum p_2 q_1 \geqslant \sum p_2 q_2$$

is a *necessary* condition for him to be as least as well off in situation 1 as in situation 2. The proof which follows makes use of the *necessary* condition rather than the *sufficient* condition, and is in the nature of a *reductio ad absurdum*.

Assume first that $$\sum P_2 Q_2 > \sum P_2 Q_1. \qquad (1)$$

Now consider the hypothesis that some reallocation of the quantities in situation 1 makes every individual at least as well off as in situation 2. Denote such a reallocation by omitting the suffix from Q and q. Since it is only a *reallocation* of the quantities in situation 1, we have at once that

$$\sum P_2 Q = \sum P_2 Q_1. \qquad (2)$$

A *necessary* condition for the hypothesis is that

$$\sum p_2 q \geqslant \sum p_2 q_2$$

[1] 'Evaluation of Real National Income', *O.E.P.*, Jan. 1950, pp. 7–8. Cf. also Hicks, 'The Valuation of the Social Income', *Economica*, May 1940, p. 111.

must hold for *every* individual, and hence that

$$\sum P_2 Q \geqslant \sum P_2 Q_2$$

and, by (2), that

$$\sum P_2 Q_1 \geqslant \sum P_2 Q_2,$$

which is inconsistent with the assumption (1).

Given, therefore, that $\sum P_2 Q_2 > \sum P_2 Q_1$, there can be no reallocation of the quantities in situation 1 which makes every individual as well off as in situation 2.

THE QUEEN'S COLLEGE,
OXFORD.

OXFORD ECONOMIC PAPERS

Volume 5. October 1953. Number 3

CONTENTS

The Full-employment Economy, with Special Reference to Wages Policy. By SIR DOUGLAS COPLAND 221

How Long does a Company Live? By GUNNAR LINDGREN . . 235

Marginal Preference and the Theory of Welfare. By JEROME ROTHENBERG 248

A Reply. By W. E. ARMSTRONG 264

The Effect of Taxation on Risk-taking. By PAUL STREETEN . . 271

Scarcity, Marxism, and Gosplan. By P. J. D. WILES . . . 288

Competition and Control in the British Sugar Industry. By P. E. HART 317

AMERICAN ECONOMIC REVIEW

CONTENTS

VOLUME XLIII DECEMBER 1953

MATHEMATICAL, or 'LINEAR,' PROGRAMMING
Robert Dorfman

VERTICAL EQUILIBRIUM—QUALITY COMPETITION
Lawrence Abbott

'REALISM' IN POLICY ESPOUSAL Clarence Philbrook

'FULL EMPLOYMENT' 1955–1960 R. S. Weinberg

Reviews of Books, Titles of New Books, Periodicals, Notes

The AMERICAN ECONOMIC REVIEW, a quarterly, is the official publication of the American Economic Association and is sent to all members. The annual dues are $6.00. Address editorial communications to Dr. Bernard .F Haley, Editor, AMERICAN ECONOMIC REVIEW, Stanford University, Room 220, Stanford, California; for information concerning other publications and activities of the Association, communicate with the Secretary-Treasurer, Dr. James Washington Bell, American Economic Association, Northwestern University, Evanston, Illinois.

Send for information booklet

SOCIAL AND ECONOMIC STUDIES

publication of

Institute of Social and Economic Research
University College of the West Indies

For socio-economic, political, strategic and other reasons the Caribbean communities are playing an increasingly important part in the affairs of the Western world. With its diversity of ethnic groups and varied local practices and traditions the Caribbean offers notable opportunities for the study of social relations, economic growth, political action and cultural change. It is these relationships which, in the main, receive attention in SOCIAL AND ECONOMIC STUDIES.

SOCIAL AND ECONOMIC STUDIES
Published: Vol. 1, Nos. 1, 2, 3, 4; Vol. 2, Nos. 1, 2, and 3
To be published from 1954 as a quarterly
Subscription per Vol. (4 nos.): £1. 10s. 0d.; U.S. $4.25; W.I. $7.20

THE ECONOMIC JOURNAL

The Quarterly Journal of the Royal Economic Society

DECEMBER 1953

A BANKER'S REFLECTIONS ON SOME ECONOMIC TRENDS — Lord Brand
BUFFER STOCKS AND INTERNATIONAL COMMODITY PROBLEMS — B. C. Swerling
THE ESTIMATION OF EQUIVALENT-ADULT SCALES FROM FAMILY BUDGETS — S. J. Prais
A REVISED THEORY OF EXPECTATIONS — C. F. Carter
DISTRIBUTION OF INCOME IN CEYLON, PUERTO RICO, THE UNITED STATES AND THE UNITED KINGDOM — T. Morgan
A NOTE ON DISEQUILIBRIUM AND THE EFFECTS OF THE ELASTICITIES OF SUPPLY AND DEMAND — A. L. Wright
A NOTE ON PLANT MECHANISATION AND UTILISATION — A. Silberston

Reviews, Recent Periodicals, and New Books

MACMILLAN & CO., LTD., LONDON, W.C.2
Application for Fellowship to
Royal Economic Society, 4 Portugal Street, London, W.C. 2
Annual Subscription £1 10 0 **Life Composition £22 10 0**

Printed in Great Britain, at the University Press, Oxford, by Charles Batey, Printer to the University